Empirical Studies in Labour and Education Economics

Elena Lisauskaitė

Thesis submitted to the University of London for the degree of Doctor of Philosophy

ROYAL HOLLOWAY, UNIVERSITY OF LONDON
Department of Economics

2019

“If” by RUDYARD KIPLING

If you can keep your head when all about you
Are losing theirs and blaming it on you,
If you can trust yourself when all men doubt you,
But make allowance for their doubting too;
If you can wait and not be tired by waiting,
Or being lied about, don’t deal in lies,
Or being hated, don’t give way to hating,
And yet don’t look too good, nor talk too wise:

If you can dream–and not make dreams your master;
If you can think–and not make thoughts your aim;
If you can meet with Triumph and Disaster
And treat those two impostors just the same;
If you can bear to hear the truth you’ve spoken
Twisted by knaves to make a trap for fools,
Or watch the things you gave your life to, broken,
And stoop and build ’em up with worn-out tools:

If you can make one heap of all your winnings
And risk it on one turn of pitch-and-toss,
And lose, and start again at your beginnings
And never breathe a word about your loss;
If you can force your heart and nerve and sinew
To serve your turn long after they are gone,
And so hold on when there is nothing in you
Except the Will which says to them: ’Hold on!’

If you can talk with crowds and keep your virtue,
Or walk with Kings–nor lose the common touch,
If neither foes nor loving friends can hurt you,
If all men count with you, but none too much;
If you can fill the unforgiving minute
With sixty seconds’ worth of distance run,
Yours is the Earth and everything that’s in it,
And–which is more–you’ll be a Man, my son!

Abstract

The first chapter of this thesis, titled "Matching Efficiency and Heterogeneous Workers in the UK" explores the consequences of the Great Recession of 2008 on the labour market in the UK, particularly, the mismatch between unemployed workers and vacancies. In this paper, I study the changes in the labour market's efficiency over the period between 2001 and 2015 in the UK, and decompose various factors behind it, such as industrial labour market segmentation and characteristics of unemployed workers, using the standard aggregate matching function. I find that the UK labour market experienced a very small if any decrease in the matching efficiency during the Great Recession.

In the second chapter, I turn to education economics. Together with Ingo E. Isphording and Arnaud Chevalier, we study the effect of ethno-linguistic classroom composition in college on performance, educational choices and post-graduation migration in a setting of quasi-random assignment to undergraduate seminars at a British university. English-speaking students are unaffected by classroom composition. Non-English-speaking students benefit from a larger linguistic diversity in terms of grades, and increase their interaction with English-speaking students. The effect of initial diversity on grades persists until the final year and is not driven by differences in specialisation. Our results imply that current levels of internationalisation do not impose a threat to native education.

In the final chapter, I examine the effects of linguistic differences between university teaching assistants on students' performance and their longer-term choices. I also present the overview of the TA gender effects on students' short and longer-term outcomes. Main findings suggest that in the short-run, non-English speaking students face lower performance outcomes as a result of being taught by a TA whose native language is other than English, however, the results vastly differ in the longer-run, suggesting that having a non-English speaking TA teaching students in early years of their studies results in higher grades in their final year. These findings are not observed for English speaking students. I also find positive gender role model effects in the beginning of the studies. Females benefit from being taught by females and males perform better when taught by male TAs. This result disappears in the longer-run – the gender of teachers in first and second year of the undergraduate degree does not have any effect on final year outcomes.

Acknowledgements

I have spent nine years at Royal Holloway. What a journey that was... One of the reasons of such a lengthy stay is people I have met. The amount of support, both academic and mental that I received from these people is unbelievable. I would like to thank my advisors, Manolis Galenianos, Jesper Bagger and Arnaud Chevalier, for teaching me how to tackle some of the biggest problems I faced. Thank you for your patience and guidance. Ija Trapeznikova, thank you for all those valuable chats in your office and thank you, Jonathan Wadsworth, for being so good to me that I could not miss an opportunity to pop into your office to say hello.

I would also like to thank my co-authors, Ingo E. Isphording and Arnaud Chevalier, for showing me what it is to work with the best colleagues one can ask for. Yes, another thank you for Arnaud, because one is not enough.

I would like to thank all the people in administration office, you've been extremely helpful and a great company. Fathima, sending you a kiss.

Thank you to Nora Sadler and Celia Blanco Jimenez. Without you two, I would have been lost.

Aleksandrai, there are no words in any language to describe what our friendship means to me. Thank you.

And last, but not least, all my family. I love you. Thank you, Mum. You are my best friend. Thank you, Dad (ačiū, Tėti), for raising me the person I am. This is for you.

Contents

List of Figures

List of Tables

Chapter 1

Matching Efficiency and Heterogeneous Workers in the UK*

ELENA LISAUSKAITĖ

Abstract

The unemployment rate in the UK increased sharply from around 5% before the Great Recession to more than 8% in the second quarter of 2009, and remained persistently high for a period of over three years. An observed rightward shift in the Beveridge Curve suggests that the efficiency in matching vacancies and unemployed workers decreased during the Great Recession. This paper studies the changes in the labour market's efficiency over the period between 2001 and 2015 in the UK, and decomposes various factors behind it, such as industrial labour market segmentation and characteristics of unemployed, using the standard aggregate matching function. Differently from the findings for the US Barnichon and Figura (2015), I find that the UK labour market experienced a very small if any decrease in the matching efficiency during the Great Recession. Accounting for labour market segmentation and worker heterogeneity as well as labour market tightness does explain more residual variation than having labour market tightness on its own.

JEL codes: J6 J41 J42
Keywords: Unemployment, Mismatch, Matching Efficiency

*I would like to thank Jesper Bagger, Manolis Galenianos, Melanie Lührmann, Ija Trapeznikova and Jonathan Wadsworth for their help and suggestions in this chapter.

1.1 Introduction

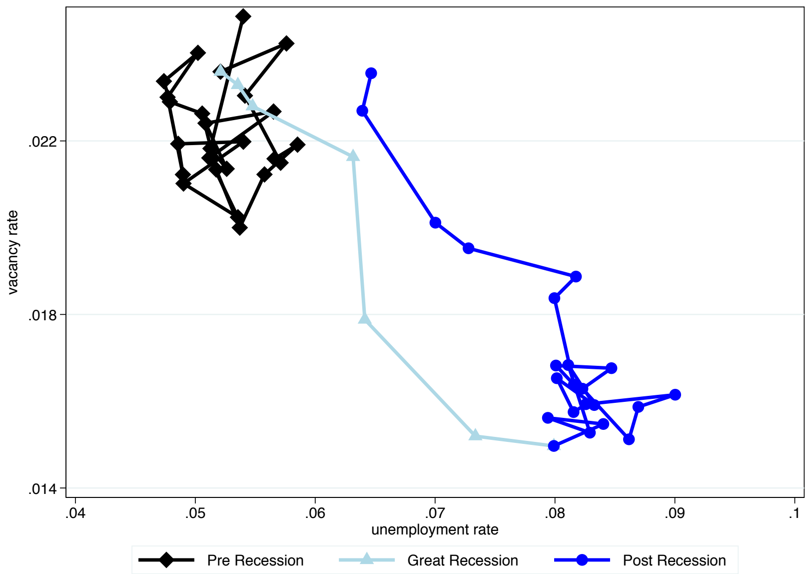


Figure 1.1: The Beveridge Curve

Source: Labour Force Survey, Vacancy Survey and author's calculations.

The unemployment rate in the UK increased sharply from around 5% before the Great Recession[1] to more than 8% in the second quarter of 2009[2]. Unemployment stayed persistently high for more than three years which has caused concern among policy makers. During the same period, number of vacancies dropped by more than 40%, however, whilst unemployment remained high, vacancies rose back to pre-recession levels relatively quickly. Other economic factors were also affected much more than they were in previous recessions of 80's and 90's. GDP fell by over 6% while real wage dropped by 4% between April 2008 and April 2011[3]. Although it took 5 years for GDP to fully recover from this recession, the UK already experienced positive output growth in 2009 Q3[4]. All these factors suggest that the UK experienced a jobless recovery from the Great Recession.

This paper focuses on the unemployment side of the labour market and the effect of the Great Recession of 2008 on it. Although the rate to which unemployment

[1] According to the OECD based Recession Indicators provided by the Federal Reserve Bank of St. Louis, the Great Recession in the UK started in November 2007 and ended in June 2009 (https://research.stlouisfed.org/fred2/series/GBRRECDM)

[2] http://www.ons.gov.uk/ons/dcp171778_424920.pdf (accessed on 27 December 2015)

[3] "Employment in the 2008-2009 recession", Economic & Labour Market Review (Vol 4, No 8, August 2010), also Blundell et al. (2014)

[4] "The 2008 recession 10 years on", ONS, 2018

increased during this recession is much lower than during the previous economic crises in the UK, it received a lot of attention from labour economists as the recovery process was particularly long and therefore, costly. The standard model of search and matching assumes a stable relation between number of matches and the fluctuation of unemployment and vacancies ratio, which is otherwise known as labour market tightness. In a frictional labour market, the matching function is assumed to be

$$m = f(U, V),$$

where m, U and V is number of matches, number of unemployed workers and vacancies, respectively. The majority of research on the effect of the Great Recession on labour market outcomes was done for the US market. It was found that one of the most important factors that prevented unemployment rate from an efficient recovery is a sharp decline in job finding probability conditional on labour demand and supply (Sahin et al. (2014)).

The main question of this paper is whether composition of unemployed workers affects the matching function in the UK over the period between 2001Q3 and 2014Q3, i.e., is there mismatch between vacancies and unemployed workers, how does it differ in different industries of the labour market, and whether changes in unemployment pool have any effect on the matching mechanism and outcomes.

Figure 1.1 plots unemployment rate against vacancy rate in the UK for the period between 2001 and 2015. This relationship is referred to as the Beveridge Curve, which typically shows a negative relationship between vacancies and unemployment (Blanchard and Diamond (1989)). The Beveridge Curve is used to distinguish between cyclical and structural changes in the labour market. The movement along the Beveridge Curve represents the cyclical changes, while shifts of the curve show structural changes in the labour market. The rightward shift of the Beveridge Curve that is observed in Figure 1.1 is the result of an increased unemployment level for a given number of vacancies. This suggests that the efficiency in matching labour market agents might have decreased during the Great Recession.

To study the questions of this paper, I employ the aggregate matching function, which takes the number of vacancies and unemployed workers as inputs and gives the number of new matches as its output. Several recent studies (e.g., (Sahin et al., 2014; Barnichon and Figura, 2015; Hall and Schulhofer-Wohl, 2018)) showed that mismatch between vacant jobs and unemployed workers arises because of a decline in average quality of unemployed workers or the fact that they tend to look for jobs in different sectors than available vacancies are. Therefore, to account for worker heterogeneity, I decompose the aggregate matching function to incorporate a number of worker characteristics, such as age, gender or the level of education. I also disaggregate the labour market into sub-markets by industries at 1-digit SIC level.

I examine the performance of the standard matching model in comparison to extended models that account for the composition of the UK labour market. That is, I estimate the aggregate matching function that includes labour market segmentation and worker heterogeneity using the UK micro data.

There is a small but rapidly growing number of papers in the literature in recent years that have focused on matching efficiency and the size of the mismatch (e.g., (Veracierto, 2011; Sahin et al., 2014; Davis et al., 2013; Barnichon and Figura, 2015)). Barnichon and Figura (2015) estimate that although the standard matching function was stable over the period from 1967 to 2007 in the US, it has broken down after 2007. After explicitly incorporating worker heterogeneity and labour market segmentation into the matching function, the authors show that the degree of heterogeneity varies substantially during recessions. They find that the two worker characteristics that are the most responsible for the break down of the standard matching function are unemployment duration and reason of unemployment. The propensity to form a match decreases as unemployment duration goes up, in addition, those who suffer a permanent job loss, are affected the most.

Sahin et al. (2013) and Sahin et al. (2014) explored the contribution of the mismatch to the rise in unemployment across different levels of disaggregation both in the US and the UK, respectively. They construct a theoretical mismatch index, which measures the fraction of new matches lost because of the misallocation of jobseekers and vacancies. Sahin et al. (2014) find that the job finding rate in 2013 was still half of what it was in 2006. The main results of both the UK and the US research suggest that there is no geographical mismatch in the labour markets, however, occupational mismatch rose steeply during the Great Recession in both countries and remained high in the UK, but declined throughout 2010 in the US.

Smith (2012) adapted the mismatch measuring model of Sahin et al. (2014) to the UK labour market. Using quarterly LFS and Vacancy Survey data, the author estimated the mismatch index and concluded that mismatch contributed to approximately one half on the increase in both steady state and actual unemployment.

As an extension of the work done by Sahin et al. (2014) and Smith (2012) on the UK, this paper contributes to the literature by providing further matching function analysis taking into account worker heterogeneity and labour market segmentation. The emphasis falls on the impact of the Great Recession on the functioning of the labour market and its ability to match unemployed workers to vacant jobs. Furthermore, this paper empirically assesses a number of unemployment and vacancy data sources available in the UK.

I show that the standard matching function, measured by OLS, slightly overpredicts the actual job finding rate, however, the magnitude of the mismatch is much smaller than the one found for the US labour market by Barnichon and Figura (2015). Accounting for differences in sub-labour markets, i.e., industries, and worker heterogeneity, does not narrow this gap between data and estimation prediction. In line with results from the US, unemployment duration is the most important component of the unemployment pool's composition effect on the job finding rate, however, I argue that this characteristic is endogenous and therefore, has to be taken out from the main estimations. I find that the sharp decline in the job finding rate in the UK after the start of the Great Recession was due to a decreased labour market tightness rather than changes in unemployment pool.

The rest of the paper is organised as follows. Section 1.2 is an overview of the data. Section 1.3 describes the methodology applied in this paper and the specification of the matching function. Section 1.4 summarises the results. Section 1.5 concludes. More detailed data description and figures are provided in Appendix A.

1.2 Data

To estimate the aggregate matching function, information about the stock of unemployed workers, the stock of available vacancies and the flow of the matches between the two is needed. In the set up of the extended version of the matching function, unemployed workers are heterogeneous, and so the estimation requires data on workers' demographics, their geographical location, and industry. The period of interest in this paper is 2001-2015, which covers a sufficient time horizon to assess the behaviour of the labour market in the UK before and after the Great Recession of 2008. Data on unemployed workers and their characteristics comes from the longitudinal quarterly Labour Force Survey (LFS). It is a 5 quarter rolling survey, i.e., respondents are followed for 5 quarters, which allows me to also use this data to form a variable of successful matches between vacancies and unemployed workers. Vacancy data is gathered from the Vacancy Survey. Finally, in this section, I will briefly talk about the alternative data sources in the UK.

1.2.1 LFS unemployment and matches

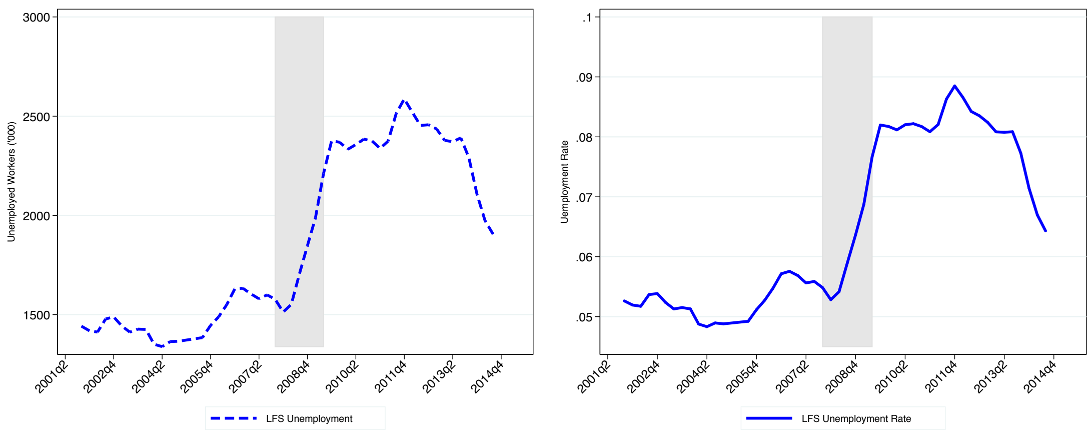


Figure 1.2: Aggregate Unemployment (LFS, Levels)

Note: Data is smoothed over 2 quarters. Shaded area represents the Great Recession as indicated by FRED (OECD).

The definition of unemployment follows the International Labour Organisation (ILO) definition. LFS allows me to calculate the number of unemployed workers for each quarter t, U_t. Furthermore, I can split the unemployed according to various characteristics that can potentially affect their job search behaviour, such as their age, gender, education level, region, ethnicity, unemployment duration, immigration status, and number of dependent children. These characteristics are taken directly from the survey data, including the unemployment duration, where people are asked how long they were unemployed, less than 3 months, 3-6 months, 6-12 months, or longer. Data is also disaggregated by industries at 1-digit SIC level. Note that the data is collected on the previous industry of unemployed workers, not the industry they are looking for a job now. In the LFS respondents are followed for five consecutive quarters and asked a number of questions about their employment circumstances. The LFS sample is made up of approximately 40,000 households and 100,000 individuals per quarter. After taking into account all sample restrictions[5], the final sample consists of 59,201 individual observations over the period between 2001Q3 and 2014Q3. For aggregate estimations of the matching function, observations are weighted by population weights provided by the LFS.

Figure 1.2 plots the unemployment series in both, levels (left panel) and rates (right panel). Unemployment increased dramatically by around 65% during the period of the Great Recession and remained this high for more than three years. Graph suggests that unemployment recovery started in 2013Q2.

Matches

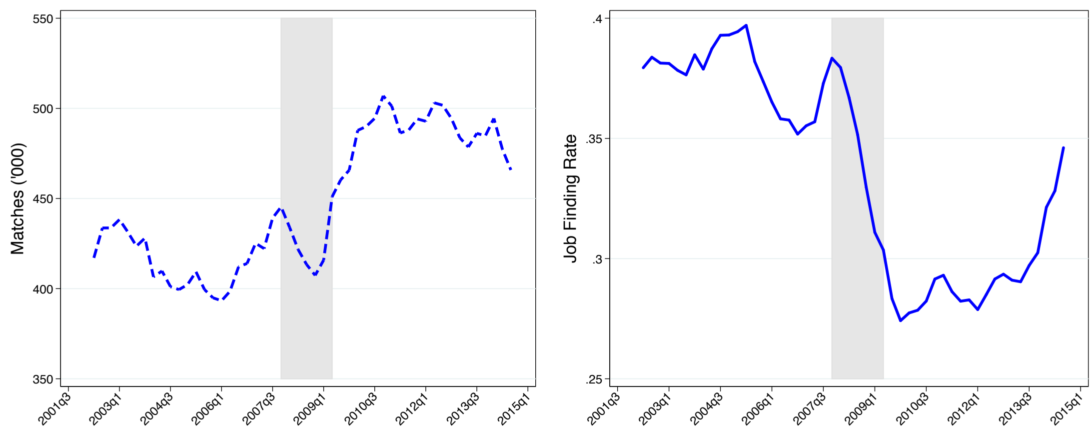


Figure 1.3: Matches and Job Finding Rate (LFS)
Note: All series are 4-quarter moving averages. Shaded area represents the Great Recession as indicated by FRED (OECD). These series are constructed from reduced population weighted sample.

A match between a vacant job and an unemployed worker in this paper is defined as a transition between unemployment state in quarter t and employment state in $t + 1$. As mentioned above, participants of the LFS are followed for five successive quarters and therefore this allows me to calculate the number of matches in each quarter. This variable is further adjusted for shorter than 1-quarter unemployment spells as people are asked how long have they been searching for a job as well as

[5]Limited number of industries surveyed by Vacancy Survey (no Agriculture, Forestry and Fishing sector); I exclude 2005Q1 as the change in employment status coding gives inconsistency in the measure of unemployment. I also need to exclude Energy & Water industry in 2006Q4 as there are too few observations and 2004Q1 as there is no information on education in this period. Missing data is then linearly interpolated

whether they changed jobs within the last three months. A job finding rate then is the ratio between the total number of new matches and the number of unemployed workers.

Figure 1.3 plots new hires (left) and the job finding rate (right). As the left panel shows, the number of matches started to sharply increase in the mid-recession. This can be explained by the sudden increase in the number of unemployed workers around the same time, thus leading to more matches. However, unemployment increased to a much greater degree than the number of successful matches, which resulted in a sharp decline in the job finding probability. There are some potential measurement issues with the job finding rate that are discussed in the Appendix.

1.2.2 Vacancies

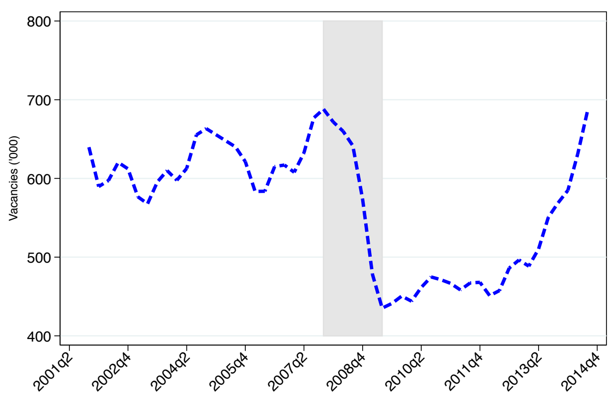


Figure 1.4: Aggregate Vacancies (Vacancy Survey, Levels)

Note: Vacancy Survey series are deseasonalised and smoothed over 2 quarters. Shaded area represents the Great Recession as indicated by FRED (OECD).

The source of vacancy data in this paper is the Vacancy Survey. It covers all industries except the Agriculture, Forestry and Fishing, which allows us to disaggregate the data into 8 sub-labour markets. The collection of the data by the Vacancy Survey started only in 2001, which will be the starting point of my estimations. The Vacancy Survey interviews 6000 businesses every month, which forms a population of 1.93 million vacancies.

As expected, the Great Recession had a sizeable impact on vacancies in the UK – the number of available jobs decreased by more than 40% (Figure 1.4). However, conversely to unemployment, vacancies began to gradually recover right after the cessation of the recession. The economic situation has led to a very gradual increase in vacancy creation in the UK with a faster increase in 2012Q2. Vacancy Survey measure of the number of available jobs reached its pre-recessional level in mid-2014, whereas unemployment rate recovered only recently[6].

1.2.3 Alternative sources of data

Data on unemployment and vacancies is scarce in the UK and matched LFS and Vacancy Survey data covers only a relatively short period of time. Different definitions of unemployment or vacancies may lead to different estimation results. Alternatively, administrative data on vacancies and unemployed workers could be used for the analysis in this paper. However, there is a number of reasons outlined below of why I trust that survey data gives more accurate and credible results.

Unemployed workers

An alternative source of unemployment data is collected by Nomis, which is a service provided by the Office for National Statistics (ONS). Nomis gives administrative data on unemployed workers who are claiming for Jobseeker's Allowance.

LFS unemployment and the claimant counts are consistent to a high degree. Both unemployment series overlap to some extent: claimants are generally recorded as unemployed under ILO definition of unemployment. However, non-claimants can appear among the unemployed if they are, for any reason, not eligible for benefits. Analogously, some people recorded in the claimant count would not be counted as unemployed. People can claim Jobseeker's Allowance if they earn low income from part-time work and therefore these people would not be unemployed.

The correlation between LFS unemployment measure and the JSA one is 0.985 before the break in the data in 2000 and 0.946 after January 2005. However, LFS measure represents the population better as it covers a longer, continuous period of time, represents wider range of professions (JSA claimants tend to seek jobs on a lower professional scale) and is adjusted by complex weights that LFS provides to represent the true magnitude of population. Due to these advantages, together with a wider choice of individual characteristics, LFS measure of unemployment is chosen as a preferred one and used for analysis in the remainder of this paper.

[6]ONS, Regional labour market statistics in the UK: April 2017

More information and detailed comparison between the two measures is given in Appendix A.

Matches

Nomis also provides two plausible measures of total matches between vacancies and unemployed workers. *Claimant off-flows*, the number of people who stop claiming Jobseeker's Allowance, is one of them. However, it is not always true that unemployed workers stop claiming benefits because they found a job. They might do so for other reasons, such as claiming benefits for a maximum period of six months, or a change in other circumstances that make claimants ineligible for JSA. Therefore, claimant off-flows, as a measure of total matches, is subject to measurement error.

Another measure of new hires in the labour market is *vacancy outflow*, which is the count of vacancies that have either been filled by JobCentre Plus or withdrawn by employers. Similarly as with claimant off-flows, we cannot assume that all vacancies were filled by unemployed workers. Some of the jobs might have been taken by people out of the labour force, or workers who experienced job-to-job transitions (without facing unemployment).

Vacancies

Similarly as with unemployment data, vacancy series are also available from two different data sources. In addition to Vacancy Survey, there is also administrative JobCentre Plus (JCP) vacancy statistics, which comes from Nomis.

JCP is the Public Employment Service for Great Britain that accounts for only about one third of the vacancies in the UK. The rest is advertised by employment agencies or directly through employers. JobCentre Plus is a nation-wide employment support service and so it is very plausible that the jobs advertised through this service are targeted at the lower end of the professions scale in terms of skill requirements. JCP vacancy data collection was discontinued in 2012.

Vacancy Survey is more representative of a real vacancy creation situation in the UK. It is not only because of a wider occupational range; in fact, no employer is obligated to notify their vacancies to Job Centres and therefore JCP measure of vacancies is generally below Vacancy Survey. In addition, both small firms and big corporations are surveyed in contrast to only the jobs notified to JobCentre Plus.

Although there are quite a few papers using administrative Nomis data to look at the matching function in the UK (e.g., (Smith, 2012; Pizzinelli and Speigner, 2017; Sahin et al., 2013)), due to reasons stated above, for the analysis in this paper, I focus on survey data rather than administrative. More detailed comparison between the two sources of data is provided in the Appendix.

1.3 Matching Function

In this section I aim to describe the three matching function specifications estimated in this paper. Starting with the standard aggregate matching function, followed by separate estimations for industry specific matching functions, and finally relaxing the assumption of homogeneous workers.

1.3.1 Aggregate matching function

The matching function is assumed to be a concave function increasing in both vacancies and unemployment

$$m_t = f(U_t, V_t),$$

where m_t is the number of new hires in a given quarter t, U_t is the stock of unemployed workers and V_t is the number of available vacancies. New matches are frequently modelled as a Cobb-Douglas matching function, which is usually assumed to exhibit constant returns to scale (CRS) (See Petrongolo and Pissarides (2001))[7]. Assuming the Cobb-Douglas form, the matching function can be written as

$$m_t = \mu_t U_t^{\sigma} V_t^{1-\sigma}, \tag{1.1}$$

where μ_t is the so called matching efficiency, which affects how quickly matches form for a given number of vacancies and unemployed workers. It consists of a constant term, μ, and an error term, ε_t, which represents random shocks to the labour market, so $\mu_t = \mu e^{\varepsilon_t}$. Empirically, the aggregate matching function can be estimated in the log-linear form

[7]In this paper, consistent with the previous findings, the assumption of CRS in the standard aggregate matching function cannot be rejected (Table 1.1)

$$\ln f_t = \ln \mu + (1 - \sigma) \ln \theta_t + \varepsilon_t, \tag{1.2}$$

where θ_t is the labour market tightness equal to $\frac{V_t}{U_t}$, and f_t is the job finding rate, $\frac{m_t}{U_t}$. $\ln \mu_t = \ln \mu + \varepsilon_t$ and therefore $\ln \mu$ is the intercept of the regression and ε_t is the error term that is assumed to be independent of explanatory variable, θ_t (i.e., strict exogeneity holds: $E(\varepsilon_t|\theta_t) = 0$). σ here is the empirical elasticity with respect to unemployment. In addition to the standard OLS estimation, the matching function is also estimated by CES and FD. To control for seasonality in the regression, a set of monthly or quarterly dummy variables is added, depending on the frequency of the data.

1.3.2 Industry specific matching function

Some of the unexplained variation in matching efficiency – the residual – might be due to industry mismatch between vacancies and unemployed workers. Different industries might have suffered different unemployment shocks due to the recession, which in turn could have affected the the cross-industry mobility. If vacancies among different industries require skills that are not transferable from industry to industry, this would result in some industries having more unemployed workers and less matches than others.

Therefore, I proceed to estimate separate matching functions for each 1-digit SIC industry under two specifications; (1) assuming that the elasticities, σ, are constant across different industries, thus the matching function can be estimated with industry fixed effects, and (2) allowing them to vary, i.e., estimating σ_i. Allowing for variation across industries, the Cobb-Douglas matching function becomes

$$m_{it} = \sum_{i=1}^{I} \mu_i U_{it}^{\sigma} V_{it}^{1-\sigma}, \tag{1.3}$$

Industry specific matching function aggregates to the standard matching function (Equation 1.1) with matching efficiency being

$$\mu_t = \sum_{i=1}^{I} \frac{U_{it}}{U_t} \mu_i \Big(\frac{\theta_{it}}{\theta_t}\Big)^{1-\sigma}, \tag{1.4}$$

Parameters of equation (1.3) then can be estimated with fixed effects using the below log-linear form

$$\ln f_{it} = \ln \mu + (1-\sigma) \ln \theta_{it} + \ln \alpha_i + \varepsilon_{it}, \tag{1.5}$$

where i is an industry and $i \in \{i, ..., I\}$, $f_{it} = \frac{m_{it}}{U_{it}}$, $\theta_{it} = \frac{V_{it}}{U_{it}}$, and α_i is the unobserved time-invariant industry effect.

To allow for the variation in the elasticities across industries (specification (2)), the aggregate matching function can be estimated industry-by-industry

$$\ln f_{it} = \ln \mu_i + (1-\sigma_i) \ln \theta_{it} + \varepsilon_{it}, \tag{1.6}$$

Estimating the matching function equation-by-equation assumes that the error terms are uncorrelated. To remove this assumption, the system of equations is also estimated using Seemingly Unrelated Regressions (SUR).

1.3.3 A matching function with heterogeneous workers

I further predict that some of the residual from the above estimations can be due to the composition of the unemployment pool. For example, someone with college degree should have higher job finding rate than a school dropout. However, the gap between these two job finding rates can change and fluctuate over time. Barnichon and Figura (2015) developed a method to explicitly incorporate worker heterogeneity into the matching function. In their matching function, unemployed workers are assumed to have different job search efficiencies depending on their type, where in this paper worker type is defined by their age, gender, education level, ethnicity, location, number of dependent children, and immigration status.

Assuming a constant elasticity across segments and including worker search efficiency, the matching function becomes

$$m_{it} = \mu_i V_{it}^{1-\sigma}(s_{it}U_{it})^{\sigma} \tag{1.7}$$

This matching function aggregates to the standard matching function 1.1 where the matching efficiency includes both industry and worker heterogeneity effects,

$$\mu_t = \sum_{i=1}^{I} \frac{U_{it}}{U_t} \mu_i s_{it}^{\sigma} \Big(\frac{\theta_{it}}{\theta_t}\Big)^{1-\sigma}, \tag{1.8}$$

where s_{it} is worker's search efficiency in sector i over period t and is equal to a weighted average of search efficiencies of specific worker types within this sector, $s_{it} = \sum_{j=1}^{J} \frac{U_{jit}}{U_{it}} s_{jit}$, j denotes worker type $(\in 1, ..., J)$. One of the problems that arises while using the LFS data to estimate the aggregate matching function with worker heterogeneity is the inability to disaggregate the data into a time series by many worker types or sub-markets as many cells then contain a zero value.

s_{jit} is assumed to have the following form

$$s_{jit} = e^{\beta X_{jit}} \tag{1.9}$$

where β is a vector of coefficients for K worker characteristics (given by vector $X_{jit} = \Big[1, x_{jit}^{1}, ..., x_{jit}^{K}\Big]$.

By giving an individual job search efficiency the above form (equation 1.9) and not allowing the parameter vector β to vary over time, it is possible to allocate a search efficiency to every set of worker characteristics.

Given that the matching function takes the form in equation (1.7), the job finding rate of an individual of type j in industry i over the period t is

$$f_{jit} = \frac{s_{jit}}{s_{it}} \frac{m_{it}}{U_{it}} = \mu_i \frac{s_{jit}}{s_{it}} s_{it}^{\sigma} \theta_{it}^{1-\sigma} \tag{1.10}$$

and so the job finding probability over period t is

$$F_{jit} = 1 - e^{-\mu_i e^{\beta X_{jit}} \left(\sum_{j=1}^{J} \frac{U_{jit}}{U_{it}} e^{\beta X_{jit}} \right)^{\sigma-1} \theta_{it}^{1-\sigma}} \tag{1.11}$$

The micro longitudinal LFS together with the Vacancy Survey provides all the data that is needed to estimate equation (1.10). Parameters σ, β and μ_i can be estimated by the following log-likelihood function

$$\ell(\beta, \mu_i, \sigma) = \sum_{t=1}^{T} \sum_{i=1}^{I} \sum_{j=1}^{J_i} \Big[(1 - y_{jit}) \ln(1 - F_{jit}) + y_{jit} \ln F_{jit} \Big], \tag{1.12}$$

where $y_{jit} = 1$ if individual of type j in industry i over a period t finds a job.

1.4 Results

1.4.1 The Aggregate Matching Function

Can the evolution of unemployment be explained by the evolution of vacancies? To answer this question, all the estimations are done for the period preceding and including the recession, before 2009Q3, to see how well the model predicts the recovery of the labour market in the UK. The predicted job finding rate is generated and compared with the observed one to measure the residual – the mismatch between the number of unemployed workers and available vacancies.

Table 1.1 presents the elasticities of the aggregate matching function estimated using various specifications. The standard OLS regression gives 0.337 elasticity with respect to vacancies, which is consistent with the previous findings[8]. The test for constant returns to scale does not reject the hypothesis, therefore, for the rest of my estimations I leave the CRS assumption in place.

Even though the Cobb-Douglas specification of the matching function is widely accepted as a good representation of the labour market, it is important to check other functional forms. Although the size of the elasticity of Constant Elasticity of Substitution (CES) estimation (column 2) is consistent with the previous findings in the literature, it is not statistically significant.

[8]Pissarides (1986) has found 0.3 elasticity with respect to vacancies for the UK between 1967 and 1983 using quarterly data.

Table 1.1: The Aggregate Matching Function: Elasticities

		OLS[9]	CES	OLS (FE)
$1-\sigma$		0.337**	0.337	0.261**
		(0.053)	(0.357)	(0.050)
	within	–	–	0.4552
R^2	between	–	–	0.2793
	overall	0.8217	0.9957	0.2705
sample size		32	32	8x32

Note: Estimations done for 2001Q3-2009Q3. Robust standard errors in parentheses. CRS test was conducted for OLS estimation with p=0.33. Constant returns to scale hypothesis cannot be rejected. Column 3 gives results of fixed industry effects where regressions are weighted by the average unemployment in each industry. ** significant at the 5 percent level. * significant at the 10 percent level.

1.4.2 Allowing for variation across segments

It may be that different industries have different matching mechanisms. As mentioned in methodology, there are two ways to incorporate industry effects into the estimations. First, is to assume constant elasticity, therefore to estimate one σ for all industries (Fixed Effects). Second, this assumption can be relaxed and σ_i can be estimated for every industry (equation-by-equation or SUR).

Fixed industry effects estimation results are presented in Table 1.1, column 3. I find that the elasticity with respect to vacancies is 0.261, which is close to the elasticity from the aggregate estimation, however, shows an upward bias of the standard matching function.

The coefficients from the equation-by-equation estimations are given in Table 1.2. The significant elasticities from standard OLS estimations vary from 0.154 to 0.500 giving results that closely lie around the elasticity from the aggregate matching function. After performing Wald tests for each regression to see whether the

[9]Due to possible nonstationarity in f_t and θ_t, the aggregate matching function is estimated in first differences (FD) to overcome the spurious correlation problem. Column 2 presents the results. The elasticity with respect to vacancies is not statistically significant from zero in my estimations, suggesting that a quarter may be a significantly long enough period to eliminate this concern.

Borowczyk-Martins et al. (2013) argue that the matching function elasticities suffer from endogeneity bias. They state that random shocks to matching efficiency affect the stock of matches both directly and indirectly through the behaviour of vacancy creation in the labour market. They found that their matching function followed ARMA(3,3) process, however, the data they used is of monthly frequency. To check if the problem exists in my data, I followed their procedure of mechanically finding the right ARMA(p,q) process to eliminate this bias. I do not find any autoregression order that is statistically significant in my data, therefore I conclude that quarterly data does not suffer from endogeneity bias.

elasticities are different from 0.3, I cannot reject this hypothesis at 5% level for any of the statistically significant from zero estimates. Some sectors, such as Manufacturing or Distribution, Hotels and Restaurants have higher elasticity with respect to vacancies, suggesting that vacancy creation in these sectors would increase the number of matches by more than an increase in vacancies in, for example, Banking sector, where the elasticity with respect to vacancies is only 0.154, however, these estimates do not fall far from an aggregate matching function. Therefore, I conclude that differences in unemployment shocks in different industries are not responsible for the mismatch that we observe in the aggregate matching function. SUR estimation results are presented in column 2 of Table 1.2. The standard error for each coefficient decreases, however, that does not affect the significance of the estimates.

For consistency, i.e., having one elasticity measure, I keep FE estimation for further analysis and comparisons.

Table 1.2: Industry (1-digit SIC2007) Segmented Matching Function: Elasticities

SIC2007		OLS	SUR
	$1-\sigma$	0.286	0.286
Energy and Water		(0.181)	(0.169)
	R^2	0.0746	
	$1-\sigma$	0.500**	0.500**
Manufacturing		(0.093)	(0.087)
	R^2	0.6391	
	$1-\sigma$	0.289**	0.289**
Construction		(0.085)	(0.080)
	R^2	0.6736	
Distribution,	$1-\sigma$	0.332**	0.332**
Hotels and Restau-		(0.063)	(0.059)
rants	R^2	0.6891	
	$1-\sigma$	0.259**	0.259**
Transport		(0.047)	(0.044)
	R^2	0.5801	
	$1-\sigma$	0.154**	0.154**
Banking		(0.054)	(0.050)
	R^2	0.5536	
Public Adminis-	$1-\sigma$	0.235**	0.235**
tration, Education		(0.119)	(0.111)
and Health	R^2	0.5734	
	$1-\sigma$	-0.138	-0.138
Other Services		(0.117)	(0.0109)
	R^2	0.2281	
	sample size	8x32	8x32

Note: Robust standard errors in parentheses. ** significant at the 5 percent level. * significant at the 10 percent level.

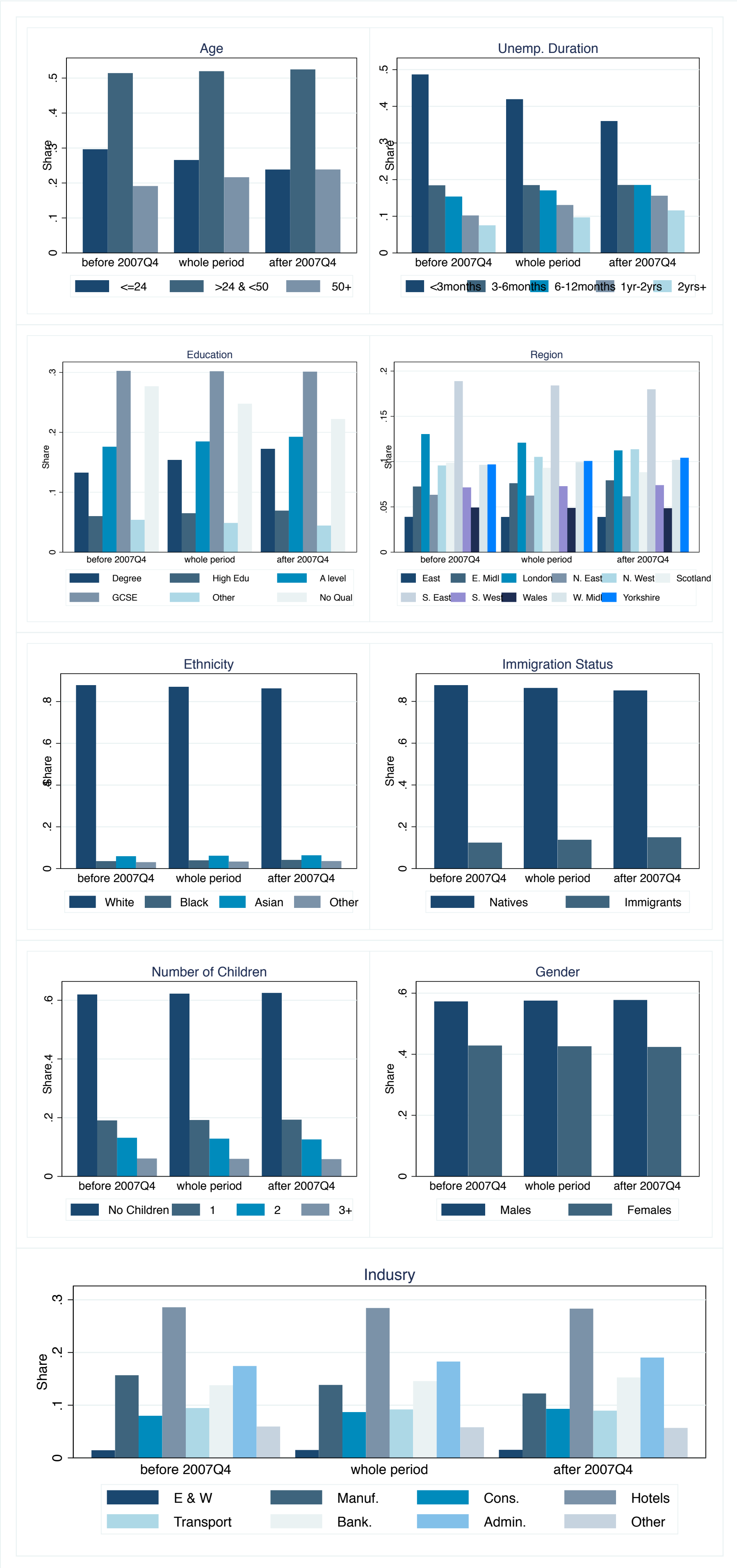


Figure 1.5: Worker composition: shares over time

Source: The Labour Force Survey (2001Q3 - 2014Q3) and author's calculations.

1.4.3 A matching function with heterogeneous workers

As previously discussed, the matching efficiency may be affected by the changing composition of workers over time. Figure 1.5 shows that indeed, there were some shifts in certain groups of unemployed workers before and after the start of the Great Recession. After the Great Recession commenced, the pool of unemployed job seekers consisted of more long term, older and better educated unemployed workers than before the recession. Another visible change was at regional level where the share of unemployed in Northern regions increased, while in Southern areas it decreased. One possible reason for such differences in unemployment rates between these regions is likely due the differences in housing market in these areas. Houses in the northern regions are up to 3.5 times cheaper than in the South, however, the pay, and especially during the recession, does not compensate for that gap[10]. Therefore, even if there were more jobs available in the South, there were mobility issues that potential workers were facing.

To estimate the matching function which simultaneously includes labour market segmentation and worker heterogeneity, the individual level LFS data is used. Given that we know if a person found a job or not during a given quarter having a certain set of characteristics (vector X_{jit}) and searching in a sector with a specific labour market tightness (θ_{it}), equation (1.11) can be estimated by maximising log-likelihood function (1.12).

Estimation coefficients: job search efficiency

One of the main reasons of estimating the matching function using individual data is to see how efficiently each of the studied group of unemployed workers is searching. Figure 6 gives the estimates of betas – the coefficients on each of worker characteristics that I am using in the aggregate ML estimations.

The results show that age plays a very important role in determining search efficiency. Consistent with Barnichon and Figura (2015) findings for the US, search efficiency is decreasing with age. Unemployed job-seekers that are 24 years old or younger are around 13.5% more likely to find a job than 50 year olds.

Non-white unemployed workers are less likely to find a job than white ones, where balck people have almost 3% lower probability to find a job within a quarter

[10]see https://www.bankofengland.co.uk/knowledgebank/how-does-the-housing-market-affect-the-economy

than white unemployed workers.. People with no qualification have up to 7% lower probability of finding a job than those with a degree.

Unemployed workers in the South have higher job search efficiency than those in North regions or Scotland. Also, higher efficiency is faced by those who have less than three dependent children in their households. Given the estimates, immigrants are around 1% more likely to find a job as well as females in comparison to males.

Table 1.3: The Matching Function (ML): Elasticities

	OLS	ML (1)	ML (2)	ML (3)
$1 - \sigma$	0.337**	0.327**	0.298**	0.302**
	(0.053)	(0.020)	(0.033)	(0.033)
Log-likelihood		-23534	-23408	-22393
Sample size	32	35,806	35,806	35,806
Quarter dummies	Yes	Yes	Yes	Yes
Industries	–	–	Yes	Yes
Worker type	–	–	–	Yes

Note: Estimations done for the period preceding and including the Great Recession, before 2009Q3. ** significant at the 5 percent level. * significant at the 10 percent level.

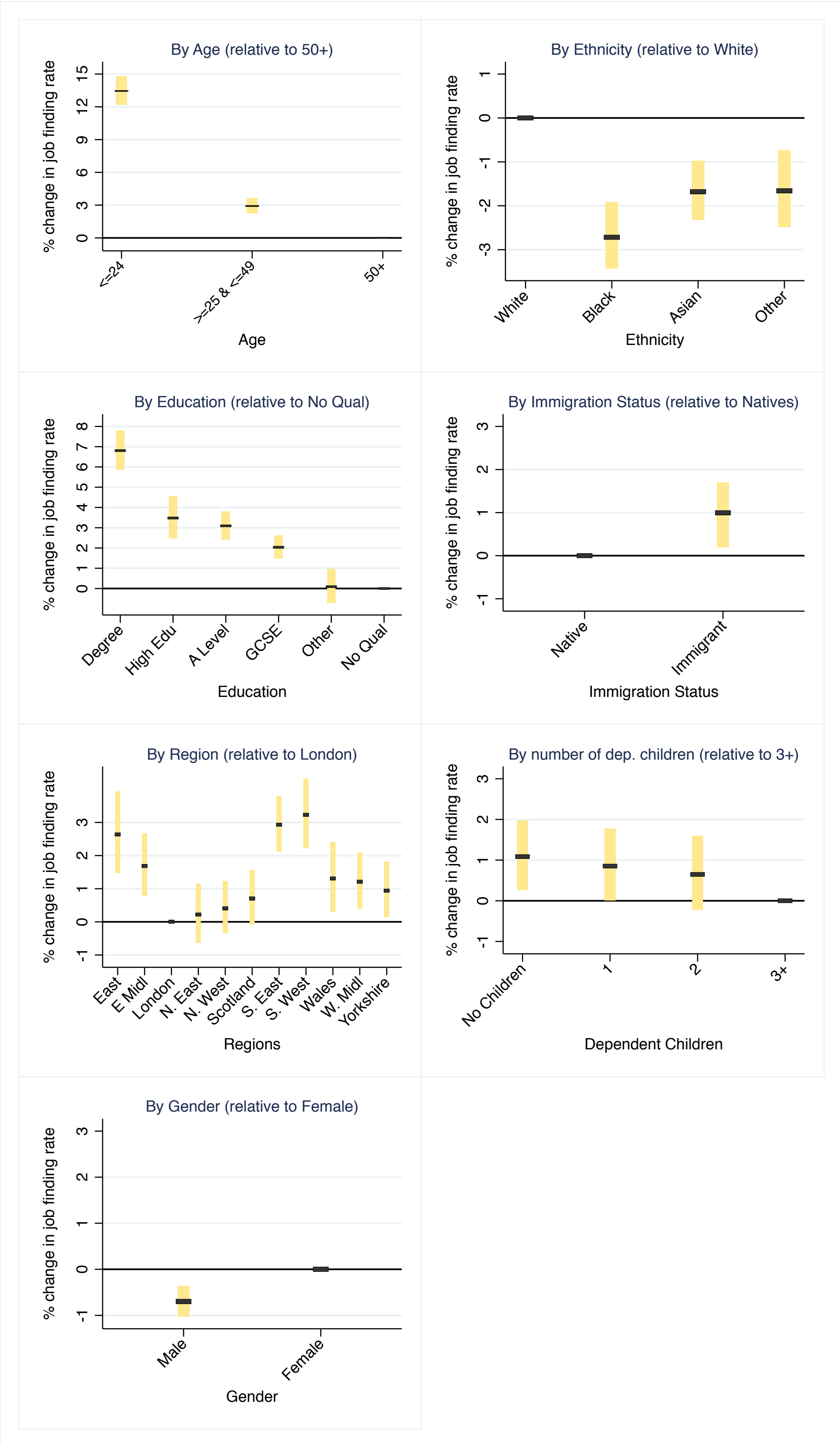


Figure 1.6: Job Search Efficiency: β

Note: Job search efficiencies from ML estimation controlling for worker composition effect. Estimated on data before 2009Q3.

Elasticities

Table 1.3 gives the comparison of the estimation coefficients from the OLS regression done on the aggregate matching function with the results from ML regressions. All the elasticities with respect to vacancies from the whole period regressions are very similar and consistent. After adding both labour market segmentation and worker characteristics to the model, the elasticity drops to 0.302 in comparison to the aggregate OLS coefficient – 0.337, which suggests that not controlling for worker heterogeneity and differences in labour market segments biases the estimates upwards. This is consistent with what Barnichon and Figura (2015) find for the US, however, their identified bias is much larger.

1.4.4 Job finding rate and movements in matching efficiency

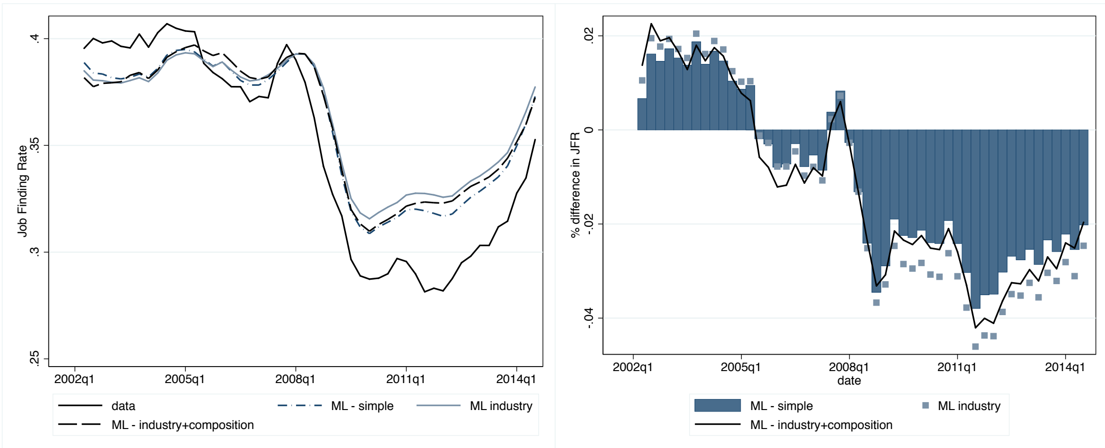


Figure 1.7: Job Finding Rates: ML estimation

Note: Predicted job finding rates and residuals from ML estimations before 2009Q3.

Figure 1.7 plots the predicted job finding rates (left) and residuals (right) from 3 estimations: (1) simple ML; (2) ML with industry effects; (3) ML including industries and worker composition to account for worker heterogeneity in the labour market. Even though standard OLS estimation of the matching function does a pretty good job at predicting the post-recession job finding rate, it seems to constantly over-predict it and so accounting for worker heterogeneity might close the gap and disclose some unexplained movements in the matching efficiency. Which is found not to be true (Table 1.4). From the results of augmented Dickey Fuller test, I can conclude that none of the residuals are following a random path. In addition, both OLS and ML residuals are consistently negative after the Great Recession. This could be explained by a decrease in GDP growth that the UK economy experienced in the second quarter of 2011[11] when the growth rate dropped below zero first time after the Great Recession and remained low until mid-2012.

Accounting for industry and worker composition effects, does not reduce the residual sum of squares (Table 1.4). In fact, the aggregate matching function, estimated by OLS, produces the lowest residuals.

Table 1.4: Residual Sum of Squares

	ML (1)	ML (2)	ML (3)
2001Q3 - 2009Q2	0.0187	0.0280	0.0223
2009Q3 - 2014Q3	0.0135	0.0151	0.0133
2001Q3 - 2014Q3	0.0322	0.0431	0.0356
Quarter dummies	Yes	Yes	Yes
Industries	–	Yes	Yes
Worker type	–	–	Yes

Note: Residual sum of squares from 2001Q3 - 2009Q3 estimations.

1.4.5 Counterfactuals: isolating the effect of composition vs. labour market tightness

It is now clear that both industry and worker composition effects do not play an important role in explaining the sharp decrease in the job finding rate at the beginning of the Great Recession. To confirm this, I segregate the composition effect from the labour market tightness, I isolate these two to see how much of the movement in the job finding rate can be explained by allowing one component to vary and restricting the other at the pre-recessional mean.

Keeping the labour market tightness constant at the pre-recessional level, the estimated job finding rate is

$$\hat{f}_t = \sum_{i=1}^{I} \mu_i s_{it}^{\sigma} \overline{\theta^{1-\sigma}},$$

where $\overline{\theta^{1-\sigma}} = \frac{1}{T_{<2007Q4}} \sum_{t=1}^{T} \sum_{i=1}^{I} \theta_{it}^{1-\sigma}$. This way, the movements coming from the labour

[11] http://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/ihyn (accessed on 01 May 2016).

market tightness are restricted and the effect on the job finding rate comes only from the changes in job search efficiency, s_{it}.

Letting the labour market tightness move freely and restricting the job search efficiency to its pre-recessional level, the job finding rate becomes

$$\hat{f}_t = \sum_{i=1}^{I} \mu_i \overline{s^\sigma} \theta_{it}^{1-\sigma},$$

where $\overline{s^\sigma} = \frac{1}{T_{<2007Q4}} \sum_{t=1}^{T} \sum_{i=1}^{I} s_{it}^{\sigma}$.

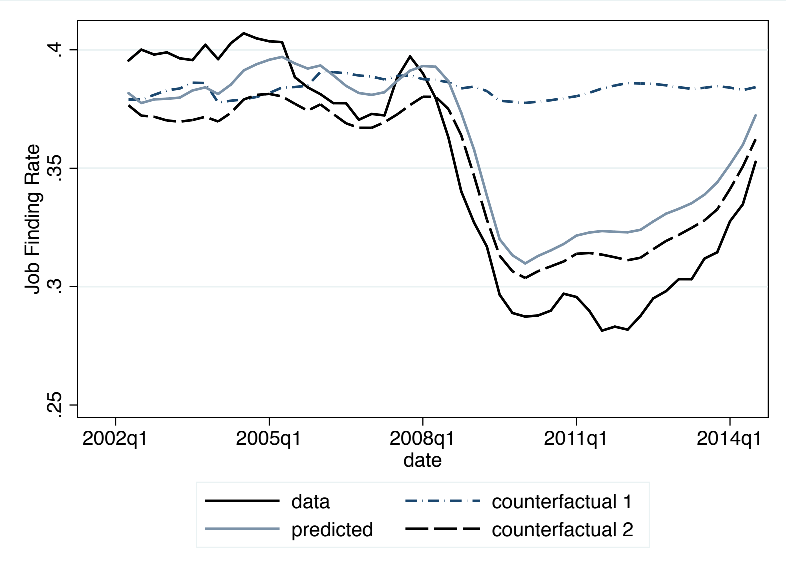


Figure 1.8: Counterfactual Job Finding Rates

Note: Predicted and counterfactual job finding rates from ML estimation of the matching function accounting for industries and worker heterogeneity. Estimation period - 2001Q3-2009Q3.

The resulting job finding rates are presented in Figure 1.8. It is easy to see that labour market tightness accounts for the movements in the job finding rate almost entirely on its own (counterfactual 2). The effect of worker composition is mainly flat and does not explain any of the drop in the job finding rate after the Great Recession (counterfactual 1). It seems that during the recession, the drop in the job finding rate was due to a plummeted labour market tightness and only in the period before the recession the composition of workers was somehow explanatory in the difference between the actual data and estimations. Therefore, the decline in the matching efficiency was mainly due to a large drop in labour demand and an increase in the number of unemployed workers rather than a worsening of the characteristics of the unemployment pool.

1.4.6 Robustness check: worker composition with unemployment duration

Barnichon and Figura (2015) base their results of composition effect's importance heavily on the duration of unemployment. In line with their results, in my estimations, if unemployment duration is accounted for, composition effect seems to play a much bigger role in explaining the decrease in the job finding rate. However, the inclusion of this characteristic into the estimations almost certainly leads to an endogeneity bias.

The lower the rate at which people find jobs, the longer the unemployment duration, d_t. Therefore, $f_t = \frac{1}{d_t}$. Trying to explain the movements of the job finding rate by the unemployment duration almost surely will cause imprecision in the estimated coefficients and possible overestimation of the true composition effect.

Nevertheless, figure 1.9 gives predicted job finding rates and residuals from the ML estimations where a set of worker characteristics includes unemployment duration. The composition effect then increases substantially and can explain up to 24% of remaining residuals. This is consistent with Barnichon and Figura's (2015) findings, however, the endogeneity bias is clear and must be accounted for.

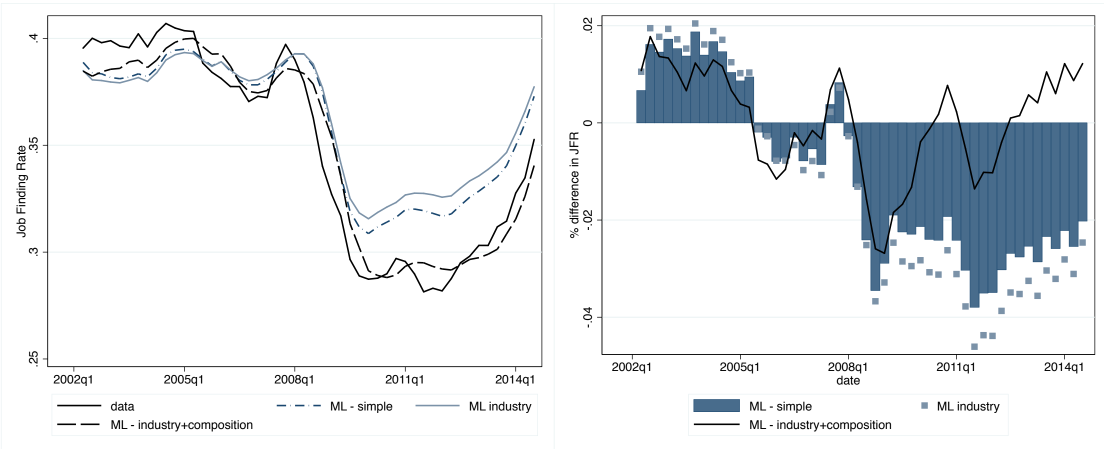


Figure 1.9: Job Finding Rates: ML estimation with unemployment duration
Note: Predicted job finding rates and residuals from ML estimations on 2001Q3-2014Q3 period.

Figure 1.10 shows the counterfactual predictions. Keeping labour market tightness constant at its pre-recessional level and allowing the composition effect to move freely (counterfactual 1) shows that now composition effect explains a significant share of the movements in the job finding rate during and after the Great Recession.

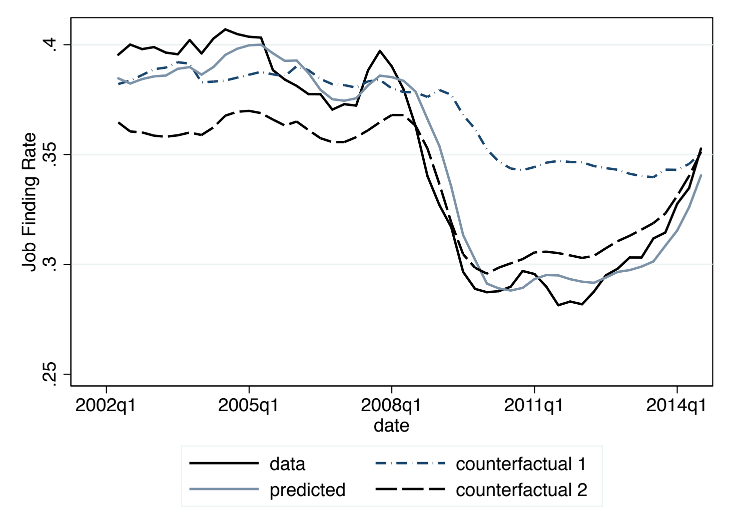


Figure 1.10: Counterfactual Job Finding Rates with unemployment duration
Note: Predicted and counterfactual job finding rates from ML estimation on 2007Q3 - 2014Q3 period.

1.5 Conclusion

Recent studies on the matching function in the UK revealed that the matching efficiency between labour market agents decreased during the Great Recession of 2008. Spurred by the lack of investigation into the reasons behind this decline in the efficiency of matching available jobs to unemployed workers, this paper extends the standard aggregate matching function for the UK to explicitly incorporate worker heterogeneity also allowing for a varying matching efficiency across different sub-labour markets – industries.

I show that consistently with the findings for the US (Barnichon and Figura (2015)), job search efficiency declines with age and the length of the unemployment. However, I argue that the inclusion of unemployment duration into worker characteristics leads to a significant endogeneity bias and therefore should be excluded from the estimation of the matching function. I show that how tight the labour markets are is the most important aspect in explaining the movements in the job finding rate. Therefore, the focus of policy makers should be on the creation of new vacancies rather than targeting specific groups of unemployed workers.

Chapter 2

Peer Diversity, College Performance and Educational Choices[◇]

ELENA LISAUSKAITĖ

co-authored with Arnaud Chevalier and Ingo E. Isphording

Abstract

We study the effect of ethno-linguistic classroom composition in college on educational performance, educational choices and post-graduation migration in a setting of quasi-random assignment to undergraduate seminars at a British university. We focus on two core variables: the share of non-English-speaking students but also the language diversity within the group of non-English-speaking students. English-speaking students are largely unaffected by the ethno-linguistic classroom composition. Non-English-speaking students benefit from a larger diversity in their grades, contemporaneously and in the long-run. Survey evidence points to increased interactions with English-speaking students as a likely mechanism. The long-run grade effects are not driven by different educational choices. Post-graduation, non-English students who have been assigned to higher shares of non-English students in the compulsory stage are more likely to leave the country. Our results imply that current levels of internationalisation do not impose a threat to native education. Avoiding segregation along ethnic lines is key in providing education for an internationalised studentship.

JEL codes: I21, I24, J15
Keywords: higher education, diversity, peer effects, foreign students

[◇]We would like to thank Benjamin Elsner, Jonas Radbruch, Ulf Zölitz, the participants of the Workshop on Higher Education at the University of Essex, the Workshop on the Economics of Education and Expectations at RHUL, the 3rd IZA Workshop on the Economics of Education, 18th IZA Transatlantic Meetings, and the European Association of Labour Economists 2019 in Uppsala. Excellent research assistance was provided by Maximilian Maehr and Youpeng Zhang. A. Chevalier would also like to thank the Nuffield foundation (grant *EDU/42242*) for partially funding this work.

2.1 Introduction

The fast-paced internationalisation of the tertiary education system has drastically increased the number of international students from 2.0 million in 2000 to 4.1 million worldwide in 2013 (UNESCO, 2018). The UK is one of the largest recipient countries of foreign students. In 2016, it hosted more than 400,000 international students, representing 18% of the student population in the country. The increasingly diverse student population has sparked a scientific and politic debate (Migration Advisory Committee, 2018). Advocates of internationalisation argue that increasing diversity benefits both native and foreign students, although critics raise concerns about potential negative spill-overs and native college flight. Empirical evidence on the effects of increasing numbers of foreign students is mostly restricted to primary and secondary education and points to ambiguous effects across different settings.[1] Tertiary education settings have received less attention to date.[2]

In this paper, we provide causal evidence on the effects of higher internationalisation and ethnic diversity on student grades and educational choices. We measure the ethnic composition of a seminar (small-scale learning group) by the share of non-English language background (labelled non-English-speaking thereafter) but also by the Ethno-linguistic diversity within the non-English-speaking group. Ethno-linguistic diversity among non-English-speaking students is expected to influence student performance and choices as it changes incentives for English language use and assimilation, a mechanism that is straightforwardly derived from a model of language assimilation (Lazear (1999)). A student from a minority broadens the pool of potential learning partners if she engages in English communication, albeit potentially at some cost if she is not a native speaker. The smaller the pool of potential same-language learning partners, the higher the incentives to engage in English communication. This intuition directly maps into a theoretically-positive effect of diversity on classroom integration and non-native performance. To the best of our knowledge, we are the first to study the effect of ethnic seminar diversity (beyond the simple share of immigrants) on student outcomes in a tertiary education

[1]The results range between moderate negative effects on native students' performance (Ballatore et al., 2018; Brunello and Rocco, 2013; Jensen and Rasmussen, 2011; Gould et al., 2009) to zero effects (Geay et al., 2013; Ohinata and van Ours, 2013).

[2]Anelli et al. (2017) show that higher shares of foreign students in introductory maths courses reduce natives' likelihood of moving into STEM majors. Braakmann and McDonald (2018) point to ambiguous effects of diversity in UK universities depending on the level of aggregation.

setting.

We base our analysis on the administrative records of economics students at a university in the London Metropolitan area. This institutional environment provides a fitting "laboratory setting" for our analysis for two reasons. First, the student body that we analyse is characterised by a high degree of ethno-linguistic diversity. Over the 2006 to 2016 period, we observe on average a share of 56 percent of non-English-speaking students from 67 different non-English-speaking countries. This level of internationalisation is shared by other British institutions, and the institution that we analyse does not stand out in terms of selectivity or graduate earnings. Second, students of the economics programme are quasi-randomly allocated to seminars during the compulsory stage of their studies, thus exposing them to exogenously-varying ethno-linguistic compositions in the classroom.

We describe four sets of results. First, grades of English-speaking students are unaffected by the share of non-English-speaking students and the diversity of a classroom. Students from a non-English background are not penalised by higher shares of non-English-speaking students, and benefit from a higher ethno-linguistic diversity; i.e. in a more diverse classroom, the academic performance of non-English-speaking students improves, especially for low-achieving students.

Second, students change their pattern of classroom interactions across ethnicities in response to higher diversity. When diversity increases, non-English-speaking students become more likely to interact with English-speaking students while interaction patterns for English-speaking students remain unaltered. Diversity appears to counteract segregation into own-language learning groups.

Third, the initial ethno-linguistic composition of seminars has persistent effects on future grades. When exposed in their early study seminars to higher shares of non-English-speaking students and a greater language diversity final year grades of non-English-speaking students improve. These long lasting effects of initial seminar compositions are not driven by changes in course choice but likely result from the altered interaction patterns in the first year.

Finally, evidence based on a voluntary alumni survey suggests that for non-English-speaking students being assigned to a larger share of non-English-speaking students in the compulsory stage increases the likelihood of being abroad at the point of interview of the post-graduation survey, potentially through a lack of native-based social networks.

Taken together, our findings suggest that 1) even at a high level of internationalisation, there are no negative effects of exposure to non-English-speaking students in small-class teaching on the learning of native students; 2) diversity improves the learning of non-English-speaking students; and 3) the ethno-linguistic composition of seminars might affect return or onward migration decisions of foreign students.

With this paper, we contribute to the literature on the effects of the internationalisation of education. In general, existing studies have focused on the share of foreign students in primary and secondary school classrooms. Ballatore et al. (2018) exploit rules of classroom formation to identify a sizeable negative effect of additional immigrant students in Italian primary schools, which is substantially larger for foreign rather than native students. Other studies describe small negative to zero to even slight positive effects of higher shares of foreign students using variation between cohorts or classes of the same school (Gould et al., 2009; Geay et al., 2013; Ohinata and van Ours, 2013; Figlio and Özek, 2017; Frattini and Meschi, 2017; Diette and Oyelere, 2017), regions (Jensen and Rasmussen, 2011; McHenry, 2015; Hunt, forthcoming) or between countries (Brunello and Rocco, 2013). A related strand of the literature has examined natives school choice in response to immigrant inflows (Betts and Fairlie, 2003). With respect to diversity of primary/secondary school classrooms, Maestri (2017) finds a positive effect of more diverse classrooms on foreigners language acquisition, while Bredtmann and Smith (2018) report negative correlations between diversity and the social integration of immigrants in Danish primary schools.

Only few studies have estimated the effects of internationalisation in tertiary education settings. Anelli et al. (2017) show that higher shares of foreign peers in introductory maths courses reduces the probability of native students moving into STEM majors, but find no direct effect on grades. Compared to our own study, their interest is on the effect of internationalisation on major choice in a US public university. In our setting, as common in European universities, students have already chosen their major at enrolment, and as such follow more comparable educational paths. Braakmann and McDonald (2018) describe a relationship between diversity at the course level and student performance for three UK-wide cohorts and deal with potentially endogenous course choices by exploiting within-programme variation across courses. They complement their main analysis with an IV strategy relying on network effects among foreign students. Their results point to ambiguous effects of diversity, depending on the level of analysis. Machin and Murphy (2018) find no evidence for a crowding out of domestic students in response to higher influx of foreign students in UK universities.

Our paper makes three contributions to these strands of literature. First, we are the first study that examines the effect of diversity in higher education beyond the simple share of foreign students on a fine-grained seminar classroom level. These seminars display relevant peer groups with meaningful social interactions. While Braakmann and McDonald (2018) demonstrates sensitivity of the level of observation with respect to student diversity, the classroom level has only been addressed by Maestri (2017) for primary school classrooms in the Netherlands and by Bredtmann and Smith (2018) for Danish secondary schools. Second, whereas the aforementioned studies have to deal with potential selection into classrooms and programs, we base our identification on a clean natural experiment relying on the quasi-random allocation of students to small-scale seminars. The as-good-as-random assignment alleviates many concerns about potential confounders through the selection of students into seminars. Third, we provide insights into plausible mechanisms by surveying students in the field about ethnic interactions and language use. Our results have implications for education practitioners. Even in an environment where non-English-speaking students represent more than half of the students, we do not find negative effects of their share on the performance and educational choices of English-speaking students. This supports current policies of pursuing greater internationalisation in higher education and should caution against forces asking for stricter admission policies discriminating by origin. Moreover, to favour the integration of non-English-speaking students, diversifying early study environments by avoiding segregated classrooms appears to be a low-cost way to reduce inequality in educational performance.

Beyond the educational setting, our results may speak to the effect of diversity in work-related settings, where existing literature has not yet reached a consensus about the effects of diversity on productivity (Hoogendoorn and Van Praag, 2012; Kahane et al., 2013; Trax et al., 2015; Dale-Olsen and Finseraas, 2019). Here, we provide complementary findings from “academic work groups” and show that higher levels of diversity indeed provide the potential to raise productivity and knowledge production.

2.2 Data and institutional setting

2.2.1 Institutional setting

We estimate the effect of the ethno-linguistic composition of seminars based on administrative data from an economics programme at a university in the London Metropolitan area. The university ranks among the top 30 universities worldwide with respect to the share of foreign students[3]. The institution we analyse is typical of the British higher education sector in terms of its organisation. It is publicly funded, selective, and tuition fees for home students are set at the maximum specified by the regulator. Graduates from this institution earn the average graduate earnings five years after graduation (Britton et al., 2018). We focus on undergraduate students registered for any of the programmes offered by the economics department, either solely or in conjunction with other departments.

Figure B.1 describes the structure of undergraduate studies. In any given year, students take four teaching units. In our empirical analysis, we focus on the ethnic composition of seminars associated to compulsory teaching units that are taught over both Spring and Autumn terms, in either the first or second year of the undergraduate degree: “Principles of Economics”, “Quantitative Methods I”, “Quantitative Methods II”, “Microeconomics”, and “Macroeconomics”. In their third, and final year, students can choose from about 20 elective courses. We regard third-year course choice as outcome variables in Section 2.4.4. Students need to obtain at least 40 out of 100 points to pass a course and a grade point average above 40 to pass a year.

Courses comprise of weekly lectures taught by a faculty member and small-group seminars where students discuss their coursework assignments. Students are assigned on an unsystematic basis, which results in an as-good-as random allocation of peers. We discuss this property and provide balancing tests in Section 2.2.4. Seminars comprise of 26 students on average, who meet for one hour per week for the full academic year, and seminars make up for about 30 to 50 percent of the

[3]https://www.timeshighereducation.com/student/best-universities/international-student-table-2018-top-200-universities

instruction time. Attendance at seminars is compulsory and monitored. Absenteeism may lead to exclusion from exams and in the case of non-EU students to visa revocation. Following the initial allocation switching to a different group is prohibited.[4]. Seminars are taught by either the course leader or teaching assistants but in both cases, the teaching materials are developed by the course leader, and all seminar leaders receive the same instructions. Teaching assistants are usually PhD students in the economics programme and they are assigned to courses, by a faculty member, independently of the assignment of students. We later control for any effect of the seminar leader by including teacher fixed effects.

2.2.2 Sample description

Each cohort comprises about 200 students, split almost equally between native English speakers and students from non-English-speaking countries. Over the period 2006 to 2016, we observe 2,184 individual students in 341 seminars, or 8,744 student-seminar observations in the five compulsory subjects. Those students who do joint degrees, do not necessarily take all five courses, however, those who do, do so compulsorily. Our data contains information about a student's contemporary performance in compulsory courses, grades and course choices in the non-compulsory stage of their studies, post-graduation migration from a post-graduate survey as well as background characteristics (gender, age and nationality). We further administered a survey on mechanisms in a contemporary cohort.

Performance. Our first main outcomes are the final grade a student receives for each compulsory course in the first and second year of the studies. These grades are computed at the end of the year and include all of the coursework, mid-term exams and final exams. All coursework and mid-term exams are marked internally by the course leader and/or teaching assistants. Marking is undertaken anonymously. The final exam, which carries the highest weight towards the final grade, is marked independently by two internal graders and checked by an external moderator.

For our analysis of student performance, we construct three outcome variables. First, we standardise the final grade within each course per year. Courses with a final grade below 40 percent are defined as having been failed. Table 2.1 (Panel A)

[4] Following a change of administration in 2016, the no-switch policy was relaxed. We therefore disregard later cohorts for our analysis.

lists means and standard deviations of the performance variables. On average, 17% of students fail a course. At the other end of the grading scale, we observe whether a student finished the course *with honours*, i.e. with an average of 60% or above. About 40% of students in our sample have a final grade of 60 points or above.

As a measure of mid-run performance, we examine effects of first year seminar composition on third-year grades. Since third year grades are based on the different courses that students have selected, and courses have different grade distributions, we standardise third-year grades by course and cohort. We thus use the average standardised grade as a measure of third-year performance which compares third year performance among students conditional on course choice.

Educational choices. We further analyse the effect of ethno-linguistic diversity on students' choices after the compulsory stage of their studies. In their third, and final year, of their studies, students choose four or more courses out of a set of about 20 different courses, irrespective of course or grade average pre-requirements. We describe a student's realised choice set by computing summary statistics of the chosen courses. First, we compute the share of numerical courses taken by a student. A course is considered to be numerical if its content is mainly quantitative and the assessments comprise more calculations rather than essay-type questions. Second, we compute the year-wise average share of non-English-speaking students in the chosen courses in that year as a measure of segregation. In other words, we compute a measure of average popularity of the chosen courses among non-English-speaking students. Third, we compute the average leave-me-out share of fails for the course as a measure of the difficulty of the realised choice set. To avoid that own choices mechanically determine this outcome, we base the computation on leave-me-out measures leaving out the current observation to ensure that own language background and choice do not mechanically alter the outcome variables.

Table 2.1 (Panel A) describes these variables based on third-year information. The self-selection of students into courses induces considerable variation over the realised choice sets. The average share of non-English-speaking students varies between 3 and 76 percent. The share of students failing in the chosen courses ranges from 0 to 20 percent. On average, 54% of students' choices in the final year are maths-intensive courses.

Post-graduation migration. To draw conclusions about a longer-term effect of ethno-linguistic seminar composition on students? post-graduation migration, we extract data from the Destinations of Leavers from Higher Education in the United Kingdom Survey (DLHE). The DLHE surveys recent graduates of each British institution six months after their graduation.[5] Graduates are contacted by e-mail, post and telephone. Being administered within the timeframe of post-graduation job search, it is not very informative about graduate career success. However, we observe whether a student has left the country.[6] We construct a binary indicator for being abroad at the time of interview. On average, 3% of English-speaking graduates among survey respondents have left the country, compared with 16% of non-English-speaking graduates (Table 2.1, Panel A). These numbers might understate the true migration if migrants have lower response rates to the survey. We will later carefully examine the role of selective response when interpreting our results.

2.2.3 Ethno-linguistic composition of seminars

Language background by nationality. We do not have direct information on the language spoken by students. To classify the ethno-linguistic composition of seminars according to the language background of students, we therefore assign language by the nationality of a student. We classify students from an English-speaking country (where English is either the predominant or an official language) as native English speakers. For non-English-speaking countries, we assign each student the predominant language of his/her nationality. While this is straightforward in most cases, we rely on a number of sources such as fact books and language encyclopaedias to determine the predominant language in the case of multilingual countries. Figure 2.1 and Table B.1 in the appendix summarise the languages, related nationalities and number of speakers in our sample. Only 44 percent of our sample are classified as native English speakers. The largest group of non-native English-speaking students is Mandarin speakers (19 percent of the sample), followed by 5 percent Russian speakers, and 3 percent Italian speakers. Overall, our sample comprises students from 68 different language backgrounds.

[5] The DLHE is organised by the Higher Education Statistical Agency (HESA). Prior to 2011/12, HESA only surveyed UK and EU domiciled graduates but since then the coverage has expanded to all graduates, and the data used pertain to this latter period.

[6] Regulations regarding post-study length of stay after graduation have changed several times over the period of interest, but graduates can always apply for visa extensions for further studies or working.

Ethno-linguistic diversity. For each seminar in our sample, we compute the share of students originating from countries not having English as the predominant or official language which we label the *non-English speaking students*. We then compute the diversity among the group of non-native English speakers by seminar. Throughout the analysis, we use leave-me-out measures of share and diversity, where the current individual observation is kept out of the computation. To describe the diversity, we use the well-known Blau Diversity Index. The Blau index, which is directly connected to the standard Herfindahl index of concentration, has a straightforward and intuitive interpretation: it measures the probability that two randomly-drawn non-English-speaking students within a seminar group have the same language background.[7] As such, it directly maps into the conceptual framework of Lazear (1999) where incentives to interact are related to the pool of potential same language partners. The Blau index is defined as

$$D = 1 - \sum_{k=1}^{K} p_k^2$$

where p_k is the fraction of language group k speakers among the non-native English speakers, excluding the current observation.

Theoretical maximum levels of the Blau Index depend on group size and the respective maximum number of potentially-distinct languages in the seminar. We account for this property by dividing the Blau Index by its respective maximum value

$$D_{max} = 1 - \frac{1}{n}$$

where n is the number of students in the group of non-English-speaking students in a seminar. This adjusted Blau Index has values ranging from 0, complete homogeneity, to 1, maximum possible heterogeneity of the non-native English speakers group and is comparable across seminars. In Section 2.4.3, we test the robustness of our results against alternative measures of diversity based on nationality only, and by assigning predominant native languages to countries with English as the second official language. We further test the robustness of our results against alternative ways of measuring diversity, such like the absolute number of languages in the classroom and the number or share of same-language peers.

[7] For very low numbers or shares of non-English students, both variables would have a discrete support. Given that only two seminars in our sample have five or fewer non-English-speaking students, we do not see discreteness of support as a threat to our empirical approach.

Language skills and linguistic distance. International students are required to provide evidence of their level of English language proficiency, with the exception of students who graduated from international schools. The main two exams that are accepted are the Test of English as a Foreign Language (TOEFL) and the equivalent International English Language Testing System (IELTS). We only observe English proficiency scores for the 42 percent of non-English-speaking students who took the IELTS. To assess the potential heterogeneity of effects by own and average seminar language proficiency among the entire sample, we proxy language proficiency by the linguistic distance to English of a student?s origin language. Following (Isphording and Otten, 2014), we assign a distance measure from the Automatic Similarity Judgment Program (ASJP) database (Wichmann et al., 2018)[8]. This measure assesses the difference in pronunciations of a culturally-independent word list (the Swadesh list) and has been shown to be a good predictor of language skill differences between immigrants (Isphording and Otten, 2014).

In Figure 2.2, we assess how linguistic distance correlates with English proficiency for the subset of students with a recorded IELTS score. We indeed observe a strong negative correlation between language distance and IELTS score, which supports the usage of linguistic distance to English as a proxy for English proficiency among the population of incoming students.

2.2.4 Random assignment to seminars

The identification of a causal effect of ethno-linguistic seminar composition relies on the idiosyncratic nature of peer assignment to seminars. Course administrators in our setting are instructed to assign students to seminars on a purely unsystematic basis. Deviations from this unsystematic assignment should only be due to scheduling conflicts due to already-assigned seminars and lectures from parallel study programmes. Indeed, only information about study programmes is known to administrators when making the allocation, but no further student characteristics. This mechanism is not sufficient per-se to guarantee randomisation if students reallocate themselves between seminars, either formally by asking to change seminar group or informally by attending a different seminar from the one they have been assigned. For the cohort of interest, the first channel is shut down as the policy

[8]The ASJP database is publicly available at https://asjp.clld.org/

was to refuse any change of seminar that was not related to a clash with another lecture/seminar. Second only attendance to the allocated seminar was recorded and thus informal reallocation would be treated as non-attendance and could lead to exclusion. Due to this institution context we therefore assume the assignment of students to seminars to be as good as random.

Although we can rely on first-hand and in-depth institutional knowledge, in the following we provide supporting empirical evidence against students, teachers or administrators circumventing the instructed unsystematic assignment out of reasons and through ways unknown to us. We therefore assess whether the observed patterns in the data are consistent with the assumed quasi-random assignment applying two different tests for quasi-randomisation.

First, we use the original data to simulate an artificial group assignment and compare the actual distribution of peer compositions against this simulated random distribution. For this purpose, we randomly assign students to artificial seminar IDs, accounting for differences in average seminar size in different courses (the average seminar size differs across courses between 16 and 37 students). We then compute the share of non-native English speakers and the ethno-linguistic diversity in each of these artificial seminars. In Figure 2.3, the white bars show the distributions of share and diversity under simulated random assignment within courses per term and based on 1,000 permutations. The grey bars show the actual observed distribution. Both distributions are of very similar shape. Permutation-based p-values of a Wilcoxon rank sum test for equality of both distributions do not reject the null at conventional significance levels (p = .892 for share and p = .549 for diversity).[9] The observed distributions therefore are a plausible outcome of the assumed quasi-random assignment to seminars within courses.

Second, we formally test whether observable pre-determined average seminar characteristics predict left-out individual characteristics. Under random assignment, no such systematic relationship should exist. We therefore regress individual-level characteristics on leave-out-shares/averages of the same characteristic (the mean value of the variable within the peer group, not accounting for the individual

[9]To determine the permutation-based p-values, we compare single simulated draws under the null as large as the observed population with the overall simulated population based on all 1,000 draws using a rank sum test. The empirical p-values are determined as the share of simulated draws which generate a test statistic as or more extreme than the one resulting from the comparison between actual observed distribution (grey bars) with the simulation-based population (white bars) displayed in Figure 2.3.

observation itself). We account for fixed effects for courses per year as the level where the randomisation takes place. Following Guryan et al. (2009) and Caeyers and Fafchamps (2016), we additionally control for the leave-out-share/average of the respective characteristic at the course/year level. This adjustment accounts for a mechanical negative correlation between own and peer characteristics that arises even under random assignment, as individuals cannot be their own peers. The results of this test are summarised in Table 2.2. Significant correlations no longer appear as soon as we control for the level of randomisation, for day/time fixed effects and study programme to account for potential deviations from random assignment due to scheduling conflicts, as expected from the institutional set-up.

Both implemented tests confirm that the observed seminar compositions in our data are consistent with the assumed random assignment of students to seminars. We therefore conclude that administrators, students, and teachers indeed do not interfer in this process, and that we can maintain the necessary identification assumption of the quasi-random assignment of students to seminars throughout our analysis.

2.3 Empirical strategy

2.3.1 Empirical model

We measure the ethno-linguistic classroom composition by the share of non-native English speakers and the ethno-linguistic diversity among the group of non-native English speakers. We identify the causal effect of these two core variables of interest by exploiting the random assignment of students to seminars. The random assignment allows us to assume differences in ethno-linguistic composition to be unrelated to students' observed or unobserved characteristics. Irrespective of their own language background, a student can experience different shares of and diversity among non-native English speakers in their seminar. A student is not able to self-select into seminars by their composition, or to select another course, since we are focusing on compulsory first- and second-year courses.

Relying on the random assignment to seminars has advantages over alternative research designs used in the literature. (Anelli et al., 2017) base their findings on arguably idiosyncratic variation over time within courses taught by the same teacher. Braakmann and McDonald (2018) use variation across courses in the same university, or within courses across universities. As the authors of these studies point out, these approaches do not entirely safeguard against selection issues that are alleviated in our setting. Nonetheless, relying on random variation has pitfalls of its own. Angrist (2014) cautions about peer effects estimated as spurious artefacts of measurement error. Feld and Zölitz (2017) instead show that under random assignment measurement error leads to an attenuation of the effect size. These issues play little role in our setting, as the variables of interest are based on information on nationality collected from administrative data which is arguably measured with very little error. [10]

We estimate the effect of the share and diversity of non-native English speakers via

$$y_{ics} = \beta_1 share_{non-English,cs} + \beta_2 D_{cs} + X_i'\gamma + Z_{cs}'\delta + \theta_{\mathbf{c}} + \epsilon_{ics}. \tag{2.1}$$

Here, y_{ics} denotes outcomes for student i, taking course c (the subscript c denotes a specific course in a specific year) and assigned to seminar s. The main variables of interest, $share_{non-English,cs}$ and D_{cs}, are the leave-me-out share of non-English-speaking students and ethno-linguistic diversity, respectively, assigned at the level of the seminar s. We additionally control for individual student characteristics (X_i), study program, age, gender, whether they are non-native English speakers and the distance of the language of their country of birth to English. Z_{cs} is a vector of seminar-level characteristics including seminar leader fixed effects, size, time and day of the seminar and the leave-me-out share of observable student characteristics. We additionally include course$\times$ year fixed effects to capture any unobservable characteristics that would be shared by all students attending a certain course in a specific year. This is also the level at which the randomisation takes place. ϵ_{ics} is the error term. We cluster standard errors at the seminar level relying on 332 clusters. We later corroborate our inference by clustering on larger levels of aggregation and using empirical p-values based on permutation exercises in Section 2.4.3.

[10] Another alternative research design in the estimation of the effect of foreign peers relies on variation by class size caps. (Ballatore et al., 2018) uses a Maimonides-type rule of class formation to estimate the effect of increasing the share of foreigners net of additional class size effects. In the absence of class size caps, such a strategy is not applicable in our context.

2.3.2 Identifying assumption and variation

We rely on a setting with as-good-as random assignment of students to seminars. While this random assignment alleviates concerns about the self-selection of students into peer groups, our identification still relies on the assumption that peer ethnicity is, conditional on observable dimensions, unrelated to unobservable peer characteristics that affect the outcomes. We later provide empirical support for the validity of this assumption by examining the coefficient stability with respect to controlling for observable peer characteristics like gender, age, residence and prior achievement.

Under this identification assumption, sufficient co-variation between share and diversity is needed to separately identify their effect. Figure 2.4 provides a schematic description of the variation in our core variables. Seminars can differ in the share of non-native English-speaking students (black symbols). Conditional on a specific share of non-native English speakers, seminars can differ in their level of diversity among the group of non-native English speakers. Comparing the hypothetical seminar B and C, it is easy to see that with a given share of non-native English-speaking students, there can be variation in the diversity, here between 0 and 1. Moving from a group of non-native English-speaking students that is fully homogeneous (seminar B, comprising speakers from one single group) to a seminar where the group of non-native English speakers comprises many different language groups (seminar C) increases the diversity while keeping the share constant.

Figure 2.5 displays the common support in both the share of non-native English speakers and the Blau Index in our raw data (upper left panel). We observe considerable variation in the share of non-native English speakers at the seminar level, ranging from about 20 to over 80 percent. Moreover, for each given share, there is also considerable variation in the diversity measure.

The upper right panel describes the respective variation in residuals after we account for fixed effects at the course×year level and control for individual and seminar characteristics. This is the relevant variation that is used in the quasi-experimental setup. Even after controlling for these factors, a considerable amount of variation remains both in the share and the Blau index, and there appears to be no relationship between the two residualised variables. Table B.2 in the appendix summarises the residual variation left in key variables after we control for the fixed effects, student and seminar characteristics according to equation 2.1. The standard deviation of the share of non-native English speakers reduces between raw measures and residuals from 0.14 (absolute) to 0.09 (residuals), and from 0.10 (absolute) to 0.07 (residuals) for the Blau Index. Running a regression of the share of non-native English-speaking students on the diversity and controlling for fixed effects according to equation (2.1) yields an insignificant coefficient of diversity of $\beta = 0.088[0.062]$. This low partial correlation ensures that effects of the share of non-native English speakers and the diversity of the non-native English speaker group can be identified separately with sufficient precision. For the remainder of the paper, we will refer to the standard deviations in residuals when describing effect sizes.

In the lower panels of Figure 2.5 the relationship between seminar size, share of non-native English-speaking students and the diversity is depicted. Both measures of ethnic seminar composition are largely independent of seminar size and are not mechanically determined by it. For both small and large seminars, we observe sufficient variation and a comparable range of share and diversity. Nonetheless, to compare students from similarly sized seminars, we control for seminar size.

2.4 Results

We first present our main results on the effect of ethno-linguistic classroom composition on the contemporary performance of students in Section 2.4.1. We then use survey evidence to describe the potential mechanisms behind the main results in Section 2.4.2. We test for the sensitivity of our results to different definitions of diversity and corroborate our inference with non-parametric permutation exercises in Section 2.4.3. We then turn to evidence on longer-term effects on third-year grade and course choice in Section 2.4.4 and post-graduation migration decisions in Section 2.4.5.

2.4.1 Performance

We start the discussion of the results with the effects of ethno-linguistic seminar composition on the contemporary performance of students. Both the share of non-native English speakers and their diversity shape the learning environment of students. The share of non-native English-speaking students will reduce the average English proficiency level in the classroom for both English and non-English speakers, which might produce negative spill-overs. However, higher diversity is expected to increase incentives to engage in English conversation for those of non-native English language background and might lead to positive performance effects for this group.

The results of estimating equation (2.1) on contemporary grades as well as indicators for failing and receiving honours are summarised in Table 2.3. The table is organised in three panels describing the average results for all student-seminar observations (upper panel), native English speakers only (middle panel), and students from a non-native English language background (lower panel).

For the full sample, the share of non-native English speakers is not significantly correlated to contemporary grades. We do observe a marginally significant positive effect of diversity on grades (column 1). A one standard deviation higher diversity increases grades by 1.9% ($0.07^{11} * 0.275$) of a standard deviation. For the probability of failing a course, both the share of non-native English-speaking students and their diversity matter (column 2). Increasing the share of non-native English-speaking students by 10 percentage points increases the probability of failing a course by 0.84 percentage points from a sample mean of 17%. We interpret this as evidence of a small negative effect of the share of non-native English-speaking students on the academic achievements of low-performing students. Higher diversity counterweights this negative effect of the share. Holding constant the share of non-native English-speaking students, a higher diversity by one standard deviation ($sd = 0.07$) reduces the likelihood of failing by 0.9 percentage points. We do not find any significant effects of either share or diversity on finishing a course with honours, thus higher-performing students seem not to be affected by seminar composition

Altering the linguistic composition of the class room might have different effects on native and non-native English speakers. A lower average ability to speak English might reduce the amount of learning for all students but also reduces the costs of engaging in the conversation, making it easier for students with low English skills to participate. Indeed, the average effects on the total sample mask a significant degree of heterogeneity in the effects. For native English-speaking students, the share of non- native English-speaking students has no significant effect on any of the outcomes of interest and the effects of diversity are close to zero.[12]

[11] residualised standard deviation of Blau Index, see figure2.5.

[12] As the lack of effects for native English students is estimated rather imprecisely, it is useful to discuss

The effect thus appears primarily driven by non- native English-speaking students, as summarised in the lower panel. A higher share of non- native English-speaking students has a negative, albeit insignificant effect on the grades on non-native English-speaking students, but these students strongly benefit from being assigned to a more diverse seminar. Increasing diversity by one standard deviation increases grades by 3.1%. The detrimental effect of the share of non- native English speakers on the probability of failing is larger than in the total sample, but remains insignificant, potentially due to the smaller sample size. The effect is again counteracted by a strong positive effect of higher diversity on performance. A higher diversity by one standard deviation reduces the probability of failing by 1.1 percentage points. Again, we find no effect of the class room linguistic composition on the probability of achieving an honour-level grade. The positive effects of diversity therefore appear to be concentrated among the lower-achieving non- native English-speaking students.

Taken together, we find that native English-speaking students are largely insensitive to the linguistic composition of seminars. This is consistent with the moderate to zero effects of foreign students on native performance found in previous studies (Brunello and Rocco, 2013; Jensen and Rasmussen, 2011; Gould et al., 2009; Geay et al., 2013; Ohinata and van Ours, 2013). Thus, it appears that the institution studied here is not out of line with other education settings (other countries, other education level) regarding the effect of non-native students on natives' learning. Note however that the imprecision of estimates implies a certain range of negative effect sizes that would not be detectable in our setting. For non- native English speaking students, on the contrary, their performance is marginally negatively affected by the share of non- native English speakers but this effect is counterbalanced by linguistic diversity. These results, consistent with Lazear (1999), are mostly driven by low achieving students.

---

the range of effect sizes we are able to detect given the estimator variance. The minimum detectable effect (MDE) of the share of non- native English speakers lies at 4.9 percent of a standard deviation in grades. The MDE with respect to diversity indicates minimum detectable increases in performance of 4.6 percent of a standard deviation in grades in response to an increased share of non- native English-speakers by 1 sd

2.4.2 Mechanisms

We now explore potential mechanisms for the contemporary effects of ethno-linguistic classroom composition on student performance. Non- native English-speaking students are affected by larger shares of non- native English-speaking students and greater diversity, implying language-based mechanisms as the most plausible candidates. Our findings support an adaption of a model akin to Lazear (1999) where the value of investing in using the majority language is greater to a member of a small linguistic minority than one of a large minority. In our setting, minority students instead broaden their pool of potential learning partners if they engage in English communication. Communicating in English bears its own costs, among others opportunity costs rising in the number of same-language peers. Thus, conditional on the share of non- native English-speaking students, diversity is theoretically linked with greater English proficiency and performance gains of non-native English speakers.

To analyse whether student interaction is affected in a way predicted by the Lazear-model, we collected survey data on social interactions and language use of the most recent cohort of students attending the same courses as those investigated in the main analysis [13]. We used an in-class written questionnaire asking students about the frequency of educational interactions with students by language background (native or non- native English-speaking students), their English use and proficiency, and perceived quality of English in the classroom.

We obtained data from 538 student×seminar observations, 222 of them with a non-English background. The overall response rate of the survey was 51 percent. Non-response is unrelated to the seminar linguistic composition.[14] We merged survey responses to seminar composition variables of the share of non- native English-speaking students and diversity. The survey questions were mainly asked on 1-5 scales regarding the frequency of interactions or quality of language proficiency, which we standardise for our empirical exercise. [15]

Table 2.4 summarises the results of estimating equation 2.1 for different items

[13]Class-room linguistic composition has same sign effects for this cohort as the one obtained for the full analysis, though coefficients are not significant due to small sample sizes.

[14]Regressing a binary indicator of non-response on the share of non- native English-speaking students and diversity using equation 2.1 yields small and insignificant coefficients of $\beta_{share} = 0.15[0.22]$ and $\beta_{diversity} = 0.02[0.31]$.

[15]The questions are displayed in Table B.9 in the appendix.

on language use and interaction by ethno-linguistic background as outcomes. The results support language use as an important mechanism behind the observed performance results. Non- native English-speaking students are substantially less likely to interact with English-speaking students in seminars where the share of non- native English-speaking students is larger (lower panel, columns 1) and are more likely to interact among each other (column 2). While this effect is partly a mechanical effect of the supply of potential interaction partners, we do not observe a similar pattern for the native English-speaking students (top panel). One explanation for the asymmetry of effects could be that non- native English-speaking students concentrate their interaction on few native speakers.

A larger pool of non- native English-speaking students thus seems to lead to stronger segregation in classroom interaction. This effect is mitigated by a higher diversity among the non- native English-speaking students: Facing a more diverse classroom increases the interaction of non- native English-speaking students with their native English-speaking fellow students. This pattern is in line with the idea sketched above of incentives to engage in interactions with the majority, which increase with the level of diversity. Effects on direct measures on perceived comfort in English use, own proficiency and the quality of English in the classroom remain insignificant for non- native English-speaking students. Native English-speaking students marginally perceive a lower quality of English spoken in the classroom when in a seminar with a higher share of non- native English-speaking students.

Taken together, the survey evidence supports the idea that different incentives to engage in interactions with native students are one of the main mechanisms to explain the effects of diversity on performance.

2.4.3 Robustness checks

We now examine the sensitivity of our results to change in the set of controls, alternative definitions of language and diversity, the specific role of Mandarin speakers, non-linearity and methods to compute standard errors.

Alternative specifications: Controlling for ability and correlated peer characteristics. The quasi-random assignment of students to seminars alleviates common empirical issues related to the selection of students into seminars. Nonetheless, the share of non- native English-speaking students might still be correlated with further dimensions of peer characteristics that have spill-overs on choices and performance by themselves, e.g. gender and ability. We control for leave-me-out averages and shares of those variables that we observe: age, gender, and the linguistic distance between country of birth and English. We further approximate peer ability through peer achievement.

Unfortunately, our data does not contain information on pre-university ability, such as entry test scores or high-school GPA. Instead, we compute student ability as a leave-seminar-out average of those grades a student has received in further classes. In Appendix Table B.3, we evaluate the stability of coefficients to the inclusion of this coarse measure of own ability and peer ability. Our observed patterns are robust to these additional controls. We still report no significant effects of seminar composition for native speakers, and positive effects of linguistic diversity on test scores, especially for low performers, among non-native English-speaking students. However, we acknowledge that this ability measure could be considered a bad control in case of spill-overs between courses.

We then test the stability of our estimated main effects towards further observable characteristics. Altonji et al. (2005) and Oster (2017) argue that movements of coefficients when controlling for observables are informative about selection on unobservables, too, as long as observable characteristics are a random subset of a larger set of characteristics. Table B.4 in the appendix summarises coefficients for the share of non- native English-speaking students and the diversity among them for four different specifications: only individual characteristics (1), additional seminar characteristics (2), and additionally controlling for own (3) and peer achievement (4). The comparison of specifications with and without seminar characteristics (1 vs. 2, and 3 vs. 4) is informative about the role of seminar characteristics as potential confounders. Estimated coefficients of diversity are fairly stable when controlling for additional peer characteristics, and not statistically different between specifications. As in the base model, we find that native English students are unaffected by seminar composition and that diversity but not share of non-native, has a positive effects on the grades of non-native students. This is not surprising: peer characteristics other than the diversity appear not to play a role in explaining the outcome, whereby the according R^2 values barely change between specifications.

Hence, we are unable to construct formal parameter bounds as in Altonji et al. (2005) and Oster (2017) but conclude that all peer observations that are observable in our data do not interact with the observed peer effects by linguistic background.

While this robustness check does not hint at confounding unobserved peer characteristics, we cannot entirely rule out that observed effects are picking up variation in unobserved peer characteristics uncorrelated with observed peer controls. This issue of potentially confounding but unobserved peer characteristics is common to any peer effects study that relies on natural variation in peers. Confounding peer characteristics cannot be separated from a person, and therefore cannot be (quasi-)experimentally stimulated. However, one might argue that owing to this inseparability, estimated effects are the relevant policy parameter. Finally, we argue that while shares and averages in further unobserved dimensions might be correlated with the share of non- native English-speaking students, this is much less likely to affect our main parameter of interest, namely the effect of diversity in the group of non- native English speakers.

Alternative definitions of language and diversity. In the main specifications, we define diversity along the lines of language groups. As such, foreign-born students from English-speaking countries are defined as native English-speaking students. This definition already anticipates language being a main mechanism in terms of how diversity affects student performance. Nonetheless, diversity could be defined along related but different dimensions.

In Table B.5, we replicate the main findings of Table 2.3, column (1), using alternative definitions of language group. Specifically, we test two alternative definitions. Column (1) lists the baseline results. In column (2), we deviate from the main specification by assigning the predominant language to countries with English as an official language.[16] While the general pattern remains, the positive effect of diversity on the grades of non-native speakers is substantially reduced. One potential explanation is that we now include students who speak English very well (as it is an official language in their country of birth) as part of the non- native English-speaking population; i.e. we introduce measurement error. These students are potentially less affected by a larger diversity. In column (3), we define nativity and diversity solely on the basis of nationality. Accordingly, we compare UK stu-

[16]Gambia, Kenya, Nigeria, South Africa, Trinidad & Tobago, Uganda

dents to non-UK students and disregard the language dimension; thus, estimates would mix the effect of language and of culture. As before, the pattern of estimated coefficients remains similar to our main specification. For the native-speaker group, alternative language definitions do not alter the base-line results. Taken together, patterns of results are fairly stable across different definitions of diversity and nativity, and are consistent with the effect being driven by language ability rather than culture.

We then explore alternative measures of diversity. In Table **??** column (2), we replace the Blau index by the number of languages spoken in the classroom. The number of languages is closely related to the diversity index (r = 0.414) and for all groups, the results indeed match the pattern found for the baseline results. Rather than diversity, it might be the lack of opportunity to speak own language that provides incentives to learn English. We thus replace diversity by the number of own language speaker (Column 3), the share of own speakers (Column 4) and an indicator for having at least one person in the seminar speaking own language. Note that these variables are negatively correlated with diversity; at a given share of non- native English Speakers a greater share of own language speakers involves a reduced diversity; we thus expect the estimate to flip sign compared to the baseline specification. Indeed, having a greater share (number) of own speakers in the classroom reduces the grades of non-English speaking students. As a further support for the hypothesis that the results are driven by changes in the demand for speaking language, we can see that a single own-language peer in the seminar is enough to reduce grades by 0.09 of a standard deviation.

These specifications confirm that our baseline results are not sensitive to alternative measures of diversity

The role of Chinese students. Diversity and share of non native -English speakers are driven to a large degree by Chinese Mandarin speakers who represent 19 percent of the sample. We therefore test the robustness of our results with respect to controlling for the number or share of Mandarin speakers. We further examine heterogeneity in effects between Mandarin speakers and other non- native English-speaking students.

Regressing the share of Mandarin speakers on the share of non- native English speakers and the diversity and controlling for fixed effects according to equation

(2.1) yields coefficients of $\beta_{share} = 0.28[0.023]$ and $\beta_{diversity} = -0.63[0.029]$ confirming that Mandarin speakers largely influence the linguistic composition of seminars.

Table B.7, column 1 displays the baseline results from Table 2.3. Despite the strong correlation between the share of Mandarin speakers and linguistic diversity, additionally controlling for the number or the share of Mandarin speakers in the seminar preserves the general pattern of the main results (columns 2 and 3). However the strong correlation between the share of Mandarin speakers and diversity reduces the precision of the diversity estimates that becomes insignificant. Splitting the sample between Mandarin speakers and other non-native English-speaking students (columns 5 and 6) again yields patterns similar to our main results for both groups; the coefficients for diversity are less precisely estimated and insignificant, albeit likely due to sample size issues. We conclude from the stability of patterns that although Chinese play an important role in generating our results, the underlying mechanisms are to some degree independent of their presence.

Non-linearity The data at our disposal is characterised by a large support for the share of non-native English speakers at the seminar level (ranging from 21% to 87%). This allows us to test for non-linearity in this effect, and assess whether there is a threshold at which the average level of English deteriorates to a level which makes learning less efficient. In Figure B.2 we report estimates of the effect of share on grades, when share of non- native English speaking students is approximated by dummies indicating its quartile in the overall distribution of non- native English speaking students shares. The estimates are small and not significantly different from zero for either population. Even at high level of internationalisation, the share of non- native English speakers has no effect on contemporary learning. When similarly measuring diversity, the greater the level of diversity, i.e. the more difficult it is to find an own-language speaker, the more non- native English speaking students, grades improve. This effect appears to be almost linear. Moreover, this effect is linear, which suggests that the optimal policy would be to maximise diversity, especially since there is no negative effect of greater diversity on the native English-speaking students.

Robustness of inference. In the main specifications, we allow for clustering of error terms at the seminar level. In Table B.8, we examine their robustness to alternative inference corrections. For reasons of clarity of exposition, we focus on our preferred specification of Table 2.3, column (1), which is replicated in column (1). Column (2) lists results when assuming i.i.d. error terms. Column (3) applies simple robust standard errors. In the remaining specifications, we adjust the level of clustering to the course×year level (column 4) and the year level (column 5). Standard errors of the share parameter increase with higher levels and smaller numbers of clusters. Standard errors of the diversity parameter appear to be insensitive. Our conclusions are thus unaltered by the choice of inference correction.

Finally, we assess the robustness of inference by using non-parametric permutation tests. For this purpose, we randomly assign students within courses to placebo seminar IDs and re-run the analysis. We repeat this procedure 2,000 times. Distributions of the resulting simulated coefficients in relation to the originally-estimated parameter are summarised in Figure B.3 in the appendix, focusing on the main results of Table 2.3, lower panel. The implied empirical p-values (p=0.556 and p=0.044, for share and diversity respectively) confirm the parametric significance levels.

2.4.4 Long-run Academic Effects

We now turn to the longer-run effects of ethno-linguistic seminar composition on final-year performance and course choice. Early seminar composition might have longer-run effects on future grades if language improvement, peer interactions or learning behaviour acquired in the compulsory stage continue to influence educational attainment.

Table 2.5 reports estimates of equation (2.1) for final-year grades. As the main table for contemporaneous grades, it is organised into three panels separately for all students, native and non-native English-speaking students. Since the distributions of grades differ between courses, we normalise grades for each course year, and the variable of interest is the average grade among the final year modules. We similarly define fail and honour as the share of final year courses that are failed and passed with a grade over 60%, respectively.

We observe a performance gain in third-year grades in response to having met a larger share of non- native English-speakers in the first two years. This positive effect of the share of non- native English speakers on third-year performance is similar for both type of speakers, (significant only marginally for native English speakers). As well as its contemporaneous effect on grades, diversity has a long-run positive effect on grades of non- native English speaking students. The effect is 60% of what was observed for contemporaneous grade, and only marginally significant. Again, diversity mostly affects low achieving non- native English speaking students, whose probability of failing is reduced when, in the compulsory stage of their studies, they attended seminars with a greater linguistic diversity. These effects are compounded by share of non-native speakers in compulsory seminars which reduces the probability of failing and increase the probability of getting honour.

Overall, a higher diversity does not only increase contemporary performance, but also subsequent grades in final year. This long-term effect is in line with a human capital effect: non-native English-speaking students who are exposed to a more diverse first year environment have higher incentives to invest in their English skills. This enables them to increase their performance contemporaneously but also in subsequent years, especially among low achievers. Contrary to the first year results, though, also a higher share of non- native English-speaking students in first year tutorials translates into *higher* performance in third year courses. Thus, the long-term effect of share of non-English students differs in sign from the insignificant but negative effect on contemporary first year performance. One plausible mechanism for this contrast of short vs long term effect is investigated in the literature on labour market effects of ethnic enclaves. While enclaves decrease early incentives to assimilate, they provide shelter from discrimination and networks providing valuable information for long-term success. A similar mechanism might possibly be at play here Moreover, as the non- native English speakers improve their language skills, their negative effects on the teaching environment decrease over-time. Indeed (Dale-Olsen and Finseraas, 2019) report that the productivity penalty associated with multi-lingual workforce disappears over time.

Additionally, we investigate whether final year effects could be driven by differences in the choice of modules. There are at least three reasons why educational choices might be affected. First, the initial effect on grades affects students' perceived academic ability, which might sway their choices towards more or less demanding courses. In particular, non- native English speakers might reinforce their perceived comparative advantage in more quantitative courses, while native English speakers similarly might perceive their comparative advantage in English as being even greater. Second, students exposed to more individuals from other ethnicities change their patterns of interaction, which might change their attitudes towards these ethnicities (Boisjoly et al., 2006; Carrell et al., forthcoming), making them more willing to interact with the same ethnicities in future courses. Third, having experienced a linguistically-dissonant learning environment in their compulsory stage, students might opt for courses with a more quantitative curriculum where verbal communication plays a lesser role (Anelli et al., 2017).

We thus focus on three indicators describing a students' realised choice set among third-year non-compulsory courses: share of numerical courses, popularity among non- native English speakers (measured as leave-me-out shares of non-native English-speaking students in current year) and difficulty (measured as the share of fails in the chosen courses).

The results do not indicate any influence of ethnic seminar composition on educational choices. The long-term effects on third year performance are not driven by students choosing courses of different difficulty or mathematical content. We do not find any effect of either share or diversity on the average failing rates of the chosen set of courses in the third year (difficulty), neither for English nor for non-English students. We do find, however, that non-English students exposed to more non-English students in the first year are more likely to choose courses which are typically more popular among non-English students. This effect of share on the course choice partly explains the observed segregation of non-English students in third year courses. We do not observe any effect of the share of foreigners on natives' choices. In this way, we differ from recent studies (Anelli et al., 2017) who report that exposure to foreign peers displaces native students away from STEM courses, and (Carrell et al., forthcoming) who finds that an initial greater share of black students increases the frequency of choosing a black roommate, but only if the initial black students are not low ability.

2.4.5 Post-graduation migration

We finally turn to the analysis of longer-term effects on post-graduation migration. While some foreign students see investing in education abroad as temporary migration before returning home, for others the returns to foreign education are greater if they remain in the country where they gain this education Dustmann and Glitz (2011). Since classroom linguistic composition affects language acquisition and educational attainment, it might alter the relative returns of staying in the UK. Additionally, early exposure to more or less foreign students might affect the social network a student is able to build, which in turn affects migration decisions. Among British students, a greater exposure to foreign students might also affect their preference for migration.

In Table 2.6, we examine the role of ethnic classroom composition during the compulsory stage of their study on post-education migration. These results are based on the Destinations of Leavers from Higher Education survey, an annual survey of recent graduates conducted by the Higher Education Statistical Agency. Since 2011/12 oversea students are also included in the sampling frame. The survey response rate is around 30% but considerably lower for non- native English speaking students (20%). In first instance we check whether survey participation is correlated with the initial seminar allocation. Compulsory-stage seminar characteristics do not affect the post-graduation survey participation of non- native English-speaking students. Native English speakers are less likely to respond if they are exposed to more diverse seminars, although the effect is small. Being exposed to a higher diversity by one standard deviation reduces the response probability by 1.3% from a mean of 39.3 percentage points. Nonetheless, this effect of treatment on the response rate cautions interpreting the effect on native English speakers as causal.

The results on migration decisions differ between native and non- native English-speaking students. Non- native English-speaking students who are exposed to more non- native English peers in their compulsory stage are more likely to have left the country at the point of the survey. A higher share of non- native English peers in a compulsory-stage seminar by 10 percentage points increases the probability of living abroad by 3.2%. This effect is in line with having fewer opportunities to build English-based networks when in contact with more non- native English peers. We cautiously interpret these results as suggestive evidence of an effect of exposure to non- native English-speaking students on their return or onward migration. Native English-speaking students are unaffected in their migration decisions by the ethno-linguistic composition of compulsory-stage tutorials. For both groups, greater linguistic diversity in the compulsory stage of higher education marginally reduces the probability of migrating. The effect is really small; a 1 standard deviation increase in class-room diversity reduces migration by 0.5 percentage points.

2.5 Conclusion

Using data from a UK higher education institution that quasi-randomly allocates students to small classes, we do not find significant evidence of a negative effect of having a larger share of non- native English-speaking students. In particular, native English-speaking students are unaffected by the ethno-linguistic composition of seminars. Non- native English-speaking students benefit from higher ethno-linguistic diversity in terms of their performance. The effect is linear and extends to grades in subsequent years. The effect is mostly driven by weaker students. Survey evidence implies that diversity augments the interaction among native and non- native English-speaking students. Diversity does not alter final year course choice or the decision to migrate.

Our results are informative for the design of classroom assignment processes. We do not find negative spill-overs of either the share or diversity of non- native English-speakers. Non- native English-speakers do however benefit in performance from higher levels of diversity, and this effect is increasing with diversity. Strategically avoiding an ethnically segregated early study environment by increasing classroom diversity therefore may be a low-cost way to prevent inequality in educational outcomes. Such re-assignment could for example be achieved through stratified randomisation to seminars, where students are randomised within their own language groups.

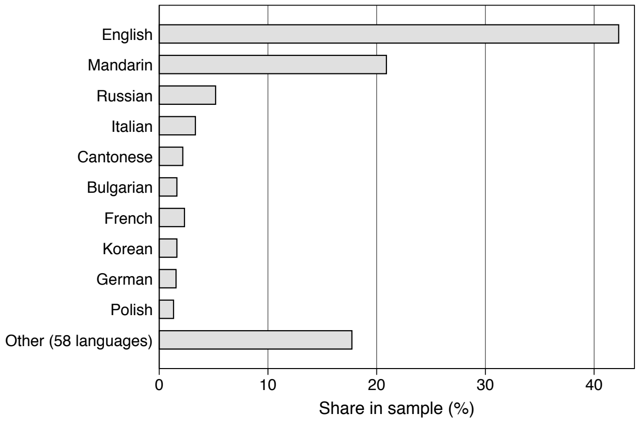
The group-work-focused learning environment in the seminars may allow for generalisations to other settings of team production involving cognitive tasks. Evidence of the effect of diversity in production settings is scarce, restricted to either quasi-experiments based on sports data (Kahane et al., 2013), lab evidence (Hoogendoorn and Van Praag, 2012) or descriptive evidence from observational data (Trax et al., 2015), (Dale-Olsen and Finseraas, 2019). Here, we add causal field evidence from a setting sharing many features of collaborative environments, which are now standard in many workplaces.

More generally, the effects of diversity on economic and social outcomes appear to differ by the level of aggregation and results are inconclusive so far. Alesina and

Ferrara (2005) propose a model that allows for negative effects of diversity on public good provision and positive effects on productivity. The majority of the literature so far has focused on the former negative effects, with evidence by Algan et al. (2016) demonstrating the negative effect of diversity on social cohesion in housing blocks in France. The latter positive effect on productivity has only recently gained attention by linking higher productivity (income) to birth place diversity (Ottaviano and Peri, 2006; Ager and Brückner, 2013). Against this broader literature on diversity, our results are informative about the positive effects of diversity on productivity on a much smaller level of peer groups with strong and meaningful social interactions.

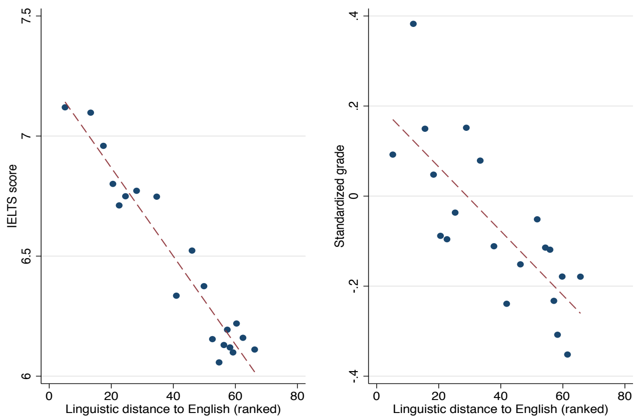
2.6 Tables & Figures

Figure 2.1: Sample composition by language background



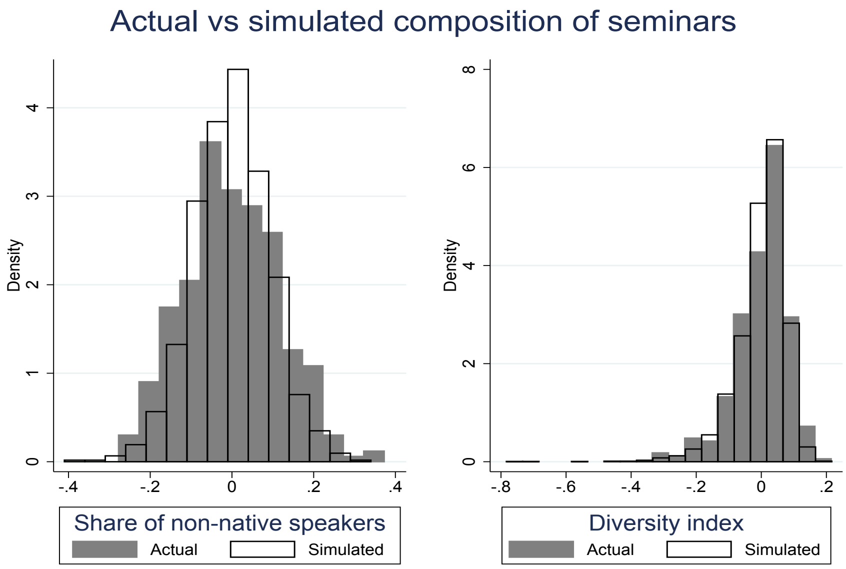
Notes: This figure displays the share of language groups in the individual sample (n=2,184). Languages are assigned by nationality: each student is assigned the predominant language of the country that the student reports as his/her nationality.

Figure 2.2: Language skills and grades by distance to English



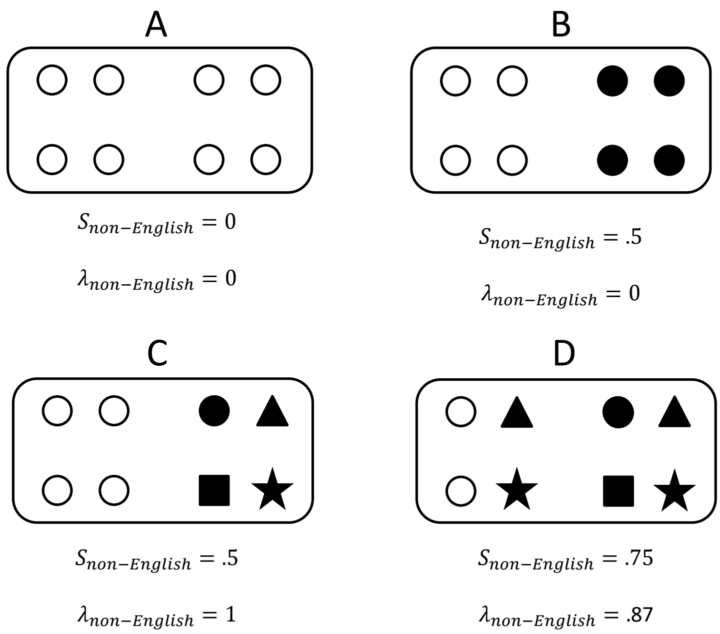
Notes: This figure displays bin scatter plots (20 bins) of the rank in linguistic distance to English to scores achieved in the *International English Language Testing System (IELTS)* (left panel, n=1,949) and the standardised course grade (right panel, n=4,712, non-English speakers only). Results displayed conditional on age, gender, linguistic distance and course×year fixed effects.

Figure 2.3: Simulated vs observed seminar composition



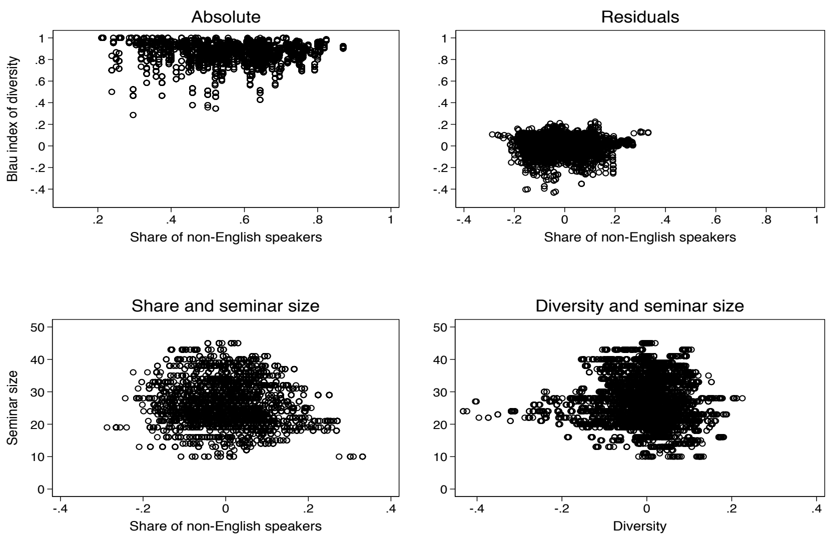
Notes: This figure compares observed distributions of the core variables of share of non-English speakers and diversity with simulated distributions based on pure random assignment based on 1,000 replications within courses, holding seminar sizes at observed levels. Variables are displayed as deviations from the *course* × *term* average. Permutation-based p-values of a Wilcoxon rank sum test cannot reject the null of equality between observed and simulated distribution.

Figure 2.4: Share vs diversity



Notes: This figure illustrates the difference between the share of non-English-speaking students and the Blau Diversity Index for classrooms with eight students. Each symbol represents a student. White dots represent English-speaking students while black symbols are for non-English speakers; each shape represents a specific native language.

Figure 2.5: Variation in share of non-English speakers and diversity



Notes: This figure displays the variation in the share of non-English speakers and the ethnic diversity within the group of non-English speakers, in absolute levels (left panel) and in residuals after applying the within-transformation by course × year, study programme, day × hour, and seminar leader fixed effects (right panel) for the Blau Index. Standard deviations: share of non-English speakers 0.14 (absolute) and 0.09 (residuals), Blau Index 0.10 (absolute) and 0.07 (residuals).

Table 2.1: Sample descriptives

	Sample: Total				Sample: English	Sample: Non-English
A. Dependent variables						
	Mean	*SD*	*Min*	*Max*	*Mean*	*Mean*
Performance						
Std. course grade	0.00	1.00	-4.75	3.14	0.08	-0.07
Course failed	0.17	0.37	0.00	1.00	0.15	0.19
With honors (above 60%)	0.40	0.49	0.00	1.00	0.43	0.38
No. of obs	*8744*				*4032*	*4712*
Educational choices						
Share of numerical courses in third year choices	0.54	0.26	0.00	1.00	0.49	0.58
Share of non-English in third year courses	0.55	0.09	0.03	0.76	0.52	0.57
Difficulty of the courses chosen in third year	0.05	0.03	0.00	0.20	0.05	0.05
Std. average grade in third year	0.00	1.00	-5.45	2.50	0.14	-0.13
Course failed in third year	0.04	0.19	0.00	1.00	0.03	0.04
With honors (above 60%) in third year	0.62	0.49	0.00	1.00	0.70	0.55
No. of obs	*7708*				*3645*	*4063*
Post-graduate outcomes						
Abroad	0.08	0.27	0.00	1.00	0.03	0.16
No. of obs	*2540*				*1583*	*957*
B. Individual characteristics						
	Mean	*SD*	*Min*	*Max*	*Mean*	*Mean*
Ling. dist. to English	51.35	47.91	0.00	104.06	0.00	95.28
Student's age	19.78	1.40	17.00	34.00	19.47	20.04
Female student	0.41	0.49	0.00	1.00	0.34	0.47
No. of obs	*8744*				*4032*	*4712*
C. Seminar characteristics						
	Mean	*SD*	*Min*	*Max*		
No. of students	*25.76*	6.83	10.00	45.00		
Share of non-English speakers	*0.55*	0.14	0.21	0.87		
Blau index of diversity	*0.87*	0.10	0.43	1.00		
No. of obs	*341*					

Notes: This table summarises descriptive statistics of individual and seminar characteristics and the dependent variables.

Table 2.2: Testing for random assignment

	(1)	(2)	(3)	(4)
Leave-me-out mean/share (seminar)				
Non-English speaker	0.376***	0.023**	0.011	0.001
	(0.052)	(0.009)	(0.010)	(0.012)
Ling. dist. to English	0.388***	0.026***	0.015	0.005
	(0.052)	(0.009)	(0.010)	(0.012)
first year GPA[a]	0.449***	0.014	0.005	-0.014
	(0.071)	(0.013)	(0.015)	(0.017)
Student's age	0.139***	-0.006	-0.007	-0.009
	(0.053)	(0.009)	(0.010)	(0.011)
Gender: Female	0.194***	0.021**	0.012	0.007
	(0.056)	(0.009)	(0.011)	(0.012)
Language: Mandarin	0.293***	0.025***	0.015	0.014
	(0.055)	(0.009)	(0.011)	(0.013)
Language: Russian	0.232***	0.016	0.016	0.014
	(0.068)	(0.013)	(0.013)	(0.015)
Language: Italian	0.132*	-0.007	-0.015	-0.020**
	(0.071)	(0.009)	(0.009)	(0.010)
Leave-me-out share/mean (urn)	yes	yes	yes	yes
Course × year FE	no	yes	yes	yes
Study program FE	no	no	yes	yes
Day/Time FE	no	no	yes	yes
Seminar leader FE	no	no	no	yes
No. of observations[a]	8744	8744	8744	8744

Notes: This table summarises results of regressions of seminar-wise leave-me-out means/shares on observable student characteristics; each row represents a separate regression. Each regression includes the course/year-wise leave-me-out mean/share and a number of fixed effects. [a]First year GPA is only available for 4,404 observations and does not generally enter our later specifications. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the individual level, are reported in parentheses.

Table 2.3: Diversity and educational performance

Sample	Grade	Fail	Honour
Total			
Share of non-English	-0.094	0.084**	0.009
	(0.129)	(0.038)	(0.059)
Blau Index	0.275**	-0.130***	-0.014
	(0.139)	(0.042)	(0.064)
Mean of dep. var.	0.000	0.167	0.401
R^2	0.05	0.08	0.13
No. of observations	8744	8744	8744
English			
Share of non-English	-0.033	0.067	-0.023
	(0.175)	(0.057)	(0.078)
Blau Index	0.004	-0.081	0.005
	(0.162)	(0.061)	(0.082)
Mean of dep. var.	0.079	0.146	0.427
R^2	0.08	0.10	0.15
No. of observations	4032	4032	4032
Non-English			
Share of non-English	-0.099	0.094*	0.040
	(0.177)	(0.056)	(0.077)
Blau Index	0.476**	-0.166***	-0.029
	(0.201)	(0.062)	(0.085)
Mean of dep. var.	-0.068	0.185	0.379
R^2	0.08	0.10	0.16
No. of observations	4712	4712	4712
Course $\times$ year FE	yes	yes	yes
Study program FE	yes	yes	yes
Day/Time FE	yes	yes	yes
Seminar leader FE	yes	yes	yes
Seminar controls	yes	yes	yes
Individual controls	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables (standardised grade, indicator for failing a course, indicator for receiving an honour (60% or above) grade) on the seminar-wise leave-me-out share of non-English speakers and the diversity index. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. Seminar controls are share of females, number of students and mean age. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table 2.4: Mechanisms

Sample	Interaction with English students	Interaction with non-Eng. students	Feeling comfortable using English	Perceived quality of English	Own English proficiency
English speakers					
Share of non-natives	-1.143	0.0952	-0.347	-1.868*	-0.101
	(0.895)	(0.864)	(0.351)	(0.990)	(0.414)
Blau Index	0.534*	-0.205	0.146	0.704	0.117
	(0.316)	(0.338)	(0.118)	(0.441)	(0.183)
Mean of dep. var.	0.28	-0.18	0.40	0.19	0.42
R^2	0.03	0.02	0.02	0.09	0.02
No. of observations	316	317	316	315	317
Non-English speakers					
Share of non-natives	-2.799**	1.947**	-0.958	-0.984	0.672
	(1.132)	(0.843)	(1.333)	(1.034)	(1.084)
Blau Index	1.454**	-1.281***	0.135	0.519	-0.500
	(0.547)	(0.386)	(0.643)	(0.462)	(0.539)
Mean of dep. var.	-0.40	0.26	-0.57	-0.27	-0.60
R^2	0.04	0.06	0.02	0.02	0.02
No. of observations	222	225	223	225	224

Notes: This table summarises results of regressions of a set of survey responses on potential mechanisms on the seminar-wise leave-me-out share of non-English speakers and the diversity index. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. Seminar controls are share of females, number of students and mean age. The survey was administered in an adjacent cohort of the autumn semester 2018. The response rate of the survey was 51 percent. Outcomes are standardised from 1-5 scales (columns 1-2: Never to Very Often, column 4: Very uncomfortable to Very comfortable, columns 4 and 6: Very bad to Very good. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table 2.5: Diversity and third-year choices

	Grade	Fail	Honour	Share of Non-English-speaking	Share of Numerical courses	Difficulty
Total						
Share of non-English	0.350***	-0.044**	0.110**	0.015**	-0.006	0.002
	(0.101)	(0.018)	(0.048)	(0.007)	(0.027)	(0.003)
Blau Index	0.219*	-0.029	0.080-0.001	0.034	0.005*	
	(0.117)	(0.022)	(0.057)	(0.007)	(0.030)	(0.003)
Mean of dep. var.	-0.000	0.036	0.621	0.549	0.537	0.051
R^2	0.10	0.04	0.11	0.68	0.33	0.63
No. of observations	7708	7708	7708	7708	7708	7708
English						
Share of non-English	0.321*	-0.021	0.038	0.005	-0.049	-0.000
	(0.190)	(0.028)	(0.075)	(0.009)	(0.037)	(0.003)
Blau Index	0.139	0.009	0.064	-0.002	-0.004	0.003
	(0.172)	(0.026)	(0.082)	(0.010)	(0.035)	(0.003)
R^2	0.11	0.08	0.14	0.63	0.33	0.68
No. of observations	3645	3645	3645	3645	3645	3645
Non-English						
Share of non-English	0.383**	-0.056**	0.171**	0.026***	0.055	0.003
	(0.152)	(0.027)	(0.079)	(0.009)	(0.037)	(0.004)
Blau Index	0.275*	-0.077**	0.137	-0.006	0.047	0.005
	(0.164)	(0.035)	(0.089)	(0.011)	(0.044)	(0.004)
R^2	0.12	0.06	0.12	0.68	0.33	0.62
No. of observations	4063	4063	4063	4063	4063	4063
Course × year FE	yes	yes	yes	yes	yes	yes
Study program FE	yes	yes	yes	yes	yes	yes
Day/Time FE	yes	yes	yes	yes	yes	yes
Seminar leader FE	yes	yes	yes	yes	yes	yes
Seminar controls	yes	yes	yes	yes	yes	yes
Individual controls	yes	yes	yes	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables regarding course choices in third year on the seminar-wise leave-me-out share of non-English speakers and the diversity index. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance and they are an English speaker or not. Seminar controls are share of females, number of students and mean age. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table 2.6: Diversity and post-graduation migration

Sample	Response	Abroad
Total		
Share of non-English	-0.040	0.060
	(0.034)	(0.040)
Blau Index	-0.066	-0.090*
	(0.054)	(0.051)
Mean of dep. var.	0.289	0.079
R^2	0.46	0.14
No. of observations	8744	2537
English		
Share of non-English	-0.029	-0.001
	(0.059)	(0.033)
Blau Index	-0.192***	-0.079*
	(0.068)	(0.041)
Mean of dep. var.	0.393	0.027
R^2	0.58	0.12
No. of observations	4032	1581
Non-English		
Share of non-English	0.012	0.323***
	(0.049)	(0.099)
Blau Index	0.037	-0.072
	(0.060)	(0.143)
Mean of dep. var.	0.203	
R^2	0.34	0.17
No. of observations	4712	956
Course × year FE	yes	yes
Study program FE	yes	yes
Day/Time FE	yes	yes
Seminar leader FE	yes	yes
Seminar controls	yes	yes
Individual controls	yes	yes

Notes: This table summarises results of regressions on response rate to a post-graduation survey and to post-graduation migratory decision on the seminar-wise leave-me-out share of non-English speakers and the diversity index. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. Seminar controls are share of females, number of students and mean age. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Chapter 3

Teacher Effects on Students' Performance and Choices in Higher Education[♮]

ELENA LISAUSKAITĖ

Abstract

Growing internationalism in the UK not only increased the share and diversity of foreign students, but also their educators. In this paper, I examine the effects of linguistic differences between university teaching assistants on students' performance in early and final years of their undergraduate degree. I also look at their longer-term choices and whether there are any effects on those from ethno-linguistic differences in TAs. Finally, I present the overview of the TA gender effects on students' short and longer-term outcomes. Main findings suggest that in the short-run, non-native English speaking students face lower performance outcomes as a result of being taught by a TA whose native language is other than English, however, the results vastly differ in the longer-run, suggesting that having a non-native English speaking TA teaching students in early years of their studies, results in higher grades in their final year. These findings are not observed for native English speaking students. I also find positive gender role model effects in the beginning of the studies. Females benefit from being taught by females and males perform better when taught by male TAs. This result disappears in the longer-run – the gender of teachers in first and second year of the undergraduate degree does not have any effect on final year outcomes.

JEL codes: I21, I23, J15, J16
Keywords: higher education, teacher effects, foreign students

[♮]I would like to thank Prof. Arnaud Chevalier and Ulf Zölitz for their help and suggestions in this chapter.

3.1 Introduction

Over the past couple of decades, tertiary education institutions saw an increase not only in the number of foreign students, but also their educators. Between years 2007 and 2014, Public Administration, Education and Health sector employment increased by approximately 9% for the UK nationals and by 27% for non-UK nationals[1]. The EU faced major enlargement in 2004 and new members also joined in 2007 and 2013. As a result, the diversity among lecturers and teachers in UK universities also increased. This is especially true for economics discipline, where according to 2016 data, 61% of academic staff is of international origin with 57% of them coming from EEA countries[2].

Given these changes in the composition of TAs in UK universities, it is important to explore the effects of such shifts on students' performance and other outcomes related to teaching staff. Even though TAs have to undergo training before they are allowed to teach, often these are postgraduate students, who might have come to the country quite recently and have not used English language often in their previous environment. Non-native English speakers also often have different accents which may be harder for students to understand and follow. These reasons spurred the research of this paper to aim and cover the question on teacher effects from the perspective of their linguistic differences.

Teacher effects have been explored by a number of papers that looked at all levels of education (Pianta (1994); White (2013); Evans (1992)), from kindergarten to university education[3], which is also the interest of this paper. There are different channels through which teachers can affect the performance of students. Role model effects have been discussed in detail, suggesting that there might be out-turns of having a teacher who has similar characteristics as the student (Fairlie et al., 2014; Lusher et al., 2018). There can also be effects of different ways of engaging with students as well as having different levels of experience.

One strand of higher education literature focused on gender effects. Carrell et al. (2010) found that female students benefit from having female professor in maths

[1] ONS and author's calculations.

[2] "World of Talent: International Staff at UK Universities & the Future Migration System" report by Campaign of Social Science, 2018.

[3] Pianta (1994) and White (2013) analysed the effects of teachers-children relationship on children's behaviour and literacy at kindergarten level. Evans (1992) looked at teacher race and gender effects in high school and found no evidence of gender role model, but positive role model effects for African-American pupils.

and science classes with minimal effects of professor's gender on males and both males and females in humanities. Hoffmann and Oreopoulos (2009) find small but positive effects of having same gender instructors in college education.

Another set of literature explored racial and ethnic interactions between teaching assistants (TAs, hereafter) and students. Fairlie et al. (2014) focused on teachers' race and ethnicity effects on students in an American college setting and found positive effects of minority teachers on the performance of minority students. Lusher et al. (2018) looked at Asian vs. non-Asian racial split and the effects of being taught by the same race TA in undergraduate courses at a university in California. They also find positive effects on Asian students' grades and future outcomes from having an Asian TA.

Borjas (2000) conducted a survey to look at the effect of foreign-born TAs on students' grades in an economics course at a large public university in the US. He found that having a foreign-born TA reduces final grades by 0.2 points. Fleisher et al. (2002) found opposing effects in another public university in the US. They conclude that foreign TAs are as good and in some cases, better (increasing grades by 0.1 points) than native TAs if they are properly trained in English, and have appropriate teaching skills.

This paper adds to the existing literature by exploring the effects of different language TAs on student performance in early years of university education as well as longer term outcomes at a non-compulsory stage of their undergraduate degree. I base my analysis on administrative data from one of the London Metropolitan area universities, focusing on economics students. In this institutional setting, students are randomly allocated to seminars (small study groups) where they are mostly taught by PhD students, who were also randomly allocated to teach different seminars. I look at five compulsory courses in the first and second year of studies and analyse the effects on academic student performance in those courses. I then look at longer term effects, performance and course choices in third, final year of the undergraduate degree.

The main variables of interest are the TA's language distance to English as well as the interaction between some biggest language groups in the data. Over the analysed period between 2006 and 2016, there were 50 TAs speaking 20 different languages. The biggest language groups were Italian and Mandarin, followed by

English-speakers (see Table 3.1). I therefore look at the effects of having a TA from one of the two biggest groups or a group of any other language in comparison to a native English-speaking TA. To see how comparable the results are with the gender literature, I finally look at the TA gender effects in this institutional setting.

I find a slightly higher probability of failing a course for non-native English speakers in the first two years of studies when they are taught by a TA with higher distance to English. However, this effect is not present for higher performing students nor is reflected in their average grades. Further on, having other than English-speaking TA, leads to an increase in the probability of failing for non-English speaking students. In addition, having a non-native English speaking TA slightly decreases non-English students' grades. Interestingly, however, having an own language TA leads to slightly lower grades. This result, however, disappears when looking at heterogeneous effects on native and non-native English speakers.

In the longer-run, the effects of having a TA whose first language is not English, leads to the opposite results. There are positive effects on both, non-English students' grades and their probability of getting honours in their final year when they are taught by non-English TAs in their first and second years. The results are consistent when looking at TAs with higher linguistic distance to English. The higher the distance, the better the performance of non-English students. Neither in short, nor in longer-term, no effects on English-speaking students are found.

Finally, I observe positive effects on grades and the probability of graduating with honours for female students when taught by female TAs. The results are somewhat consistent with what Carrell et al. (2010) find, however, this effect in my paper is short lived and disappears when looking at longer term outcomes. I also find negative effects on the same outcomes for male students when taught by female TAs, which leads to the conclusion that males also benefit from own gender TAs. This is in line with the literature of role model effects.

This paper contributes to the existing literature on the teacher effects by expanding it to language interactions between students and teachers as well as adding to - so far limited - exploration of this topic in tertiary education.

3.2 Data

3.2.1 Sample description

To analyse the effects of different TAs' languages on students' performance and longer-term outcomes, this paper uses administrative data from an economics programme at one of the universities in London Metropolitan area. I focus on five compulsory courses in the first two years of the Batchelor's of Science degree in Economics either pure or joint with other departments, excluding degrees where economics is a minor subject.

The TA effects are explored within seminars - small, compulsory and attendance monitored study groups. Given quasi-random allocation of students into seminars and seminars to TAs, controlling for the characteristics of students and seminars, I can extract the effects of linguistic differences of TAs on students' short- and long-term performance as well as their future choices.

The dataset used in this paper is the same as in Chapter 2 of this thesis, now adding TA characteristics as well. Therefore, the following data description concentrates only on the additional information. The data enables me to construct one of the main variables, TAs' language distance to English, using a distance measure from the Automatic Similarity Judgment Program (ASJP) database (Wichmann et al. (2018)). As with student data, I assign language to TAs by their nationalities[4]. TAs' linguistic distance to English ranges from 72.61 for German TAs to furthest language from English, Mandarin, with distance of 101.27. Theoretically, linguistic distance is bound between 0 (identical languages) to 100 (as dissimilar as two random sets of characters would be). A number over 100 means that languages are more dissimilar than they would be if they consisted of two randomly chosen sets of characters (Isphording and Otten (2014)). In addition to nationality and language of the TAs, the data also contains their age, gender and experience in teaching the same course.

Table 3.2 summarises the new variables. In the university setting, TAs often are PhD students, therefore they are young professionals, early in their careers, having, at most, 5 years of experience. In this dataset, all TAs are PhD students, on average, just below 29 years old with 1.5 years of experience in teaching the same course. I am

[4]For more information and references on the language distance estimation, refer to Chapter 2 of this thesis.

interested in the effects of TAs' differences on student outcomes and therefore, the following descriptives are based on the number of students TAs get to teach rather than the number of TAs. 47% of students are taught by female teachers. The biggest linguistic groups among TAs are Italian, Mandarin and English speakers, representing 17%, 14% and 11% of the sample, respectively. 9% of students are taught by their own language TA.

3.2.2 Random assignment

The identification of the causal effect of TA characteristics on students' outcomes relies on a random assignment of seminars to TAs and no self-selection to seminars neither from students nor from their teachers. The former can be ruled out by the fact that students do not have the option of selecting into seminars themselves and as explained in detail in Chapter 2, are quasi-randomly assigned to seminars.

Table 3.4 summarises the results. I account for fixed course/year effects as well as for the average of the respective seminar characteristic at the course/year level. In addition, I control for study programme and day/time fixed effects. The only significant predictors that I observe, e.g., older TAs prefer female students or Italian TAs prefer younger students, can be ruled out as it is not possible for the TAs to make such structural choices prior to seminar allocations. Given no existing correlations between and among dependant and independent variables, I can conclude that TAs are assigned randomly and identification strategy holds.

To complete the tests and not to rely on assumptions, I also look for any systematic correlations between student and TA characteristics. As shown in Table 3.3, no set of characteristics are correlated, suggesting that students do not self-select into seminars with TAs of particular characteristics. This analysis is done at student level.

3.3 Empirical Strategy

To identify the teacher effects on students' performance, I estimate the following equation

$$y_{ics} = \beta_1 TAChar_{ics} + X_i'\gamma + P_{ics}'\lambda + Z_{cs}'\delta + \theta_{\mathbf{c}} + \epsilon_{ics}. \tag{3.1}$$

Here, y_{ics} denotes outcome for student i, taking course c (the subscript c denotes a specific course in a specific year) and assigned to seminar s. $TAChar$ represents one of the independent variables of interest. These include: linguistic distance between teacher's language and English, dummy variables indicating the language TA is speaking, a dummy of whether the teacher speaks the same language as the student and finally, the gender of the TA.

I control for individual student characteristics (X_i), age, gender, whether they are non-English speakers and the distance of the language of their country of birth to English. P_{ics} is teacher related characteristics for a particular student in a particular seminar. These are seminar leader age, experience teaching the same course and gender. Z_{cs} is a vector of seminar-level characteristics including the size of the seminar and the leave-me-out share of observable characteristics, age and gender. I additionally include course×year fixed effects to capture any unobservable characteristics that would be shared by all students attending a certain course in a specific year. ϵ_{ics} is the error term. I cluster standard errors at the individual student level, which gives 2120 clusters for the full sample.

3.4 Results

I first show the results of teachers' linguistic characteristics on students' performance in the first years of their undergraduate degree. I then turn to longer-term effects and look at TAs' effects on students' performance in their third, final year. I also look at the effects on their course choices, the difficulty and the share of maths intensive courses. In Section 3.4.3, I analyse the effects of TAs' gender on students short and longer-term performance. Section 3.4.4 presents the robustness check of the main findings relaxing the ability controls.

3.4.1 Performance in early stages of university studies

In this section I analyse the effects of linguistic differences between TAs and their students on the contemporary performance; students' grades, probability of failing or finishing the year with honours in that particular course.

As argued in Chapter 2, the distance between languages measured using ASJP database, is a good proxy for language proficiency. I look at the effects of the TA language's distance to English. Table 3.5 summarises the results. I do not observe any statistically significant effects either on grades or the probability of getting honours. When looking at heterogeneous effects by native/non-native English speaking student groups, I find a significant positive effect on fail rates for non-native English students when taught by a TA with higher distance to English. A one standard deviation increase in linguistic distance to English leads to an increase in probability of failing the course by 3 percentage points[5]. This suggests that lower performing non-English speaking students are somewhat negatively affected by TAs with larger distance to English, however, the results are very close to zero.

The results of TAs' linguistic distance to English are small and not very informative. This may be because of some measurement error, i.e. linguistic distance to English can be the same to two different languages that may have different impacts on students. Other measures of differences in languages TAs speak need to be looked at.

To further explore the effects of different languages of TAs, it is worth to dissect the sample into the biggest language groups that TAs speak and look at the effects of particular languages. I look at four main categories. Italian, Mandarin and all other languages in comparison to English speaking TAs. First panel of Table 3.6 shows the effects on the whole sample of students. There is no significant effect on grades when TAs speak Italian or Mandarin, but there is an overall slightly negative effect of having a TA speak any other language rather than English. However, this effect is small and in magnitude similar to the effects of having other language TAs. Both Italian and other language speakers increase the probability of failing the course. Splitting the sample into native and non-native English speakers reveals that native English speaking students' outcomes are unaffected by having a non-English speaking TA. However, these results are very different for non-native English speaking students. They face both, lower final grades and higher probability of failing the course when taught by a TA who is not a native English speaker. Having other than English TA, reduces grades for non-English speakers by 0.05 points of grades' standard deviation, which is 17.21. Being taught by non-English TA also increases probability of failing by about 3 to 4 percentage points. This is consistent with the above findings from variation in TAs' linguistic distance to English.

[5]Here and later in this chapter, the percentage points are calculated by multiplying standard deviation of the dependent variable by the estimated coefficient. This gives clearer interpretation of the magnitude of the results.

Even though, I do observe some negative effects significantly different from zero, they are very small and I can conclude that there does not seem to be any effect on students' performance from having TAs with different linguistic backgrounds.

Finally, I analyse the effects of having your own language TA. This is shown in Table 3.7. Only a small subset of the sample, less than 9%, are taught by TAs whose mother tongue is the same as the students'. However, looking at this variable of interest, some unusual results are revealed. Increasing the probability of having a TA who is your native language speaker by one standard deviation, decreases grades by 0.027 points of grades' standard deviation when analysing the whole sample. This also marginally increases the probability of failing the course and decreases the chances of getting honours. However, all of these results disappear when looking at English/non-English student groups separately.

3.4.2 Longer-term effects

The dataset also allows me to look at longer term effects of the same TA characteristics. The outcome variables in third and final year of undergraduate degree are the share of numerical courses, difficulty of the course (leave me out share of fails in that particular course), final grade and dummies for whether the course was failed and whether honours was achieved. These variables are described in detail in Section 2.2 of Chapter 2.

Table 3.8 summarises the effects of TAs' language distance to English in first and second year seminars on third year outcomes. It is important to point out that longer term effects seem to be greatly different from the short term effects of the same TA characteristics. Despite me not finding any effects on the maths intensive courses or the difficulty of the choices, there seem to be positive effects on students' performance in their final year when taught by a TA with higher distance to English. A one standard deviation increase in distance to English leads to 0.06 points increase in grades for the whole sample, where the result is driven by the benefit to non-English speaking students. Even though there is no effect for English speakers, one standard deviation increase in the TA's distance to English leads to 0.09 points increase in non-English speaking students' grades (of one standard deviation in grades, which in final year is 16.38). Non-English speaking students also enjoy 3 percentage points higher probability of finishing their final year with honours.

Looking at the effects of different language TA groups on students' performance in their third year, I can propose the same conclusion. It appears that in the longer-run, having other than English TA is actually beneficial for non-English speaking students. Their grades increase as well as the probability of getting honours. This is presented in Table 3.9. There is no effect on the performance of English speakers. However, they do choose more difficult courses as a result of being taught by other language speaking TAs (other than the big three, Italian, Mandarin and English). Same TAs also result in the course choices of non-English speaking students containing less mathematical courses.

The final table, Table 3.10, in this section summarises the effects of having your own language TA on third year outcomes. There are no effects observed on students' performance in their final year. However, English speaking students who are taught by English TAs tend to choose less difficult courses whereas non-English speaking students, when taught by their language TAs seem to be choosing more quantitative courses.

Linguistic differences between TAs have significant impact on students' short and longer-term outcomes. Native English speaking students' performance is not affected, however, non-native English students face higher probability of failing a course in the short-term as a result of being taught by a TA (in the first two years) whose language distance to English is larger. Having other than English speaking TA also reduce non-English speaking students' grades. In the longer term, however, the same TAs result in opposite effects for the same group of students. They increase the grades and increase the probability of finishing final year courses with honours.

As English speaking students are not affected by different language speaking TAs, one plausible explanation for such results might be different language levels that students have in the beginning of the studies. Worse performance in the first years could also boost these students' motivation and in the end increase their efforts, resulting in higher overall grades in third year. In Section 3.4.4, I relax some of the controls to see what drives the results.

3.4.3 Gender effects

Earlier in the paper, I reviewed some of the existing literature on teachers' effects where the characteristic of interest was TA's gender. I was interested to see how my results compare to that literature.

Table 3.11 presents the results on short-term performance outcomes from being taught by a female TAs in the first and second year of students' degree. Consistently with the previous literature[6], I find positive effects on both grades and probability of getting honours for female students. An increase in probability of being taught by a female by one standard deviation, increases female students' grades by 0.036 of grades' standard deviation and increases probability of getting honours by 2 percentage points. The effects for males are opposite and of similar size. This shows role model effects that go both ways in the first years of students' undergraduate degree. Females benefit from being taught by females and males get higher grades and probability of receiving honours when taught by male TAs.

These positive results disappear when looking at longer term effects. There are no observed significant effects from TAs' gender in early seminars on any of the outcomes in third year. It seems to be important to have your own gender teacher in the beginning of the studies, but as students proceed and get familiar with the requirements, gender effects die out.

3.4.4 Robustness check

One of the most important open questions still remaining is the opposite short and longer-term effects on non-English students when being taught by non-English TAs in early years' seminars. As I touched upon before, given that there are no such effects on English-speaking students, this suggests that the language skills of non-English students' in the beginning of their studies play some role. However, in my analysis, I control for students' linguistic distance to English and the results are robust to the exclusion of this control.

Another plausible channel that could influence such results is students' ability. In all of the regressions, I control for student individual and leave-me-out average peer ability. I exclude these controls from both, short and longer-term regressions and present the results in Tables 3.13 and 3.14. The effects on the early perfor-

[6] see (Carrell et al., 2010; Hoffmann and Oreopoulos, 2009)

mance of non-English students decrease when I do not control for ability. Now the coefficients on grades are smaller and no longer significant and probability of failing also decreases. In the final year, the positive effect on grades and probability of getting honours is even higher if ability controls are excluded.

This suggests that the group of students that is the most affected is the high-achievers. Coming from different backgrounds, it may be difficult to understand different accents and different teaching methods. However, a slight decrease in grades in the first years, may suggest that students are still adapting to new environment and teaching methods and in turn have a positive effect on the motivation of students and efforts that they put in later in their studies. Being high-achievers in the first place, once they overcome the language and cultural barriers, their performance increases even further. Some role model effects may also be in place that could result in higher grades of non-English speakers when taught by non-English TAs that become prominent once students integrate into the new environment. Even though this analysis allows us to look into the results deeper, to understand this effect fully, one would need to look into the differences between English and non-English TAs, their teaching techniques and backgrounds.

3.5 Conclusion

After the recent increase in diversity among higher education teachers in the UK, in this paper, I study the effects of having different linguistic background TAs in university study groups on students' short and longer-term outcomes. In addition, I look at teachers' gender effects in the same setting. I use administrative data of economics students at a university in the London Metropolitan area.

The main results suggest that native English speaking students' performance is unaffected by having a foreign TA, neither in short, nor in longer-term. However, I observe significant effects for non-native English speaking students. In the beginning of their studies, being taught by a TA whose linguistic distance to English is larger, leads to a slightly higher probability of failing the course and being taught by any other language speaking TA rather than native English speaker, leads to lower overall grades as well. However, the short term results are very small and I conclude that having different linguistic background TAs does not have an effect on the short-term performance of students. This result is opposite in the longer-run. Having a non-native English speaking TA in the beginning of the studies, has a strong positive effect on final year results. Non-native English speaking students benefit from higher grades and have a higher probability of finishing the year with honours.

I argue that one of the reasons for these results to differ is the effect on higher-achievers. Due to possible and unobservable language barriers, difficulty to understand different accents that the TAs are speaking, these students are doing badly in the beginning, but due to integration and motivation, in the final year, they do even better.

In this paper I also present evidence of gender role model effects in the short term. Both females and males benefit from being taught by same gender TA in the first two years of their studies. However, this result is short lived and does not appear in the third year results.

Given the empirical findings in this study, one may suggest to introduce better techniques of foreign students' integration as well as to further study the fundamental differences between English and non-English speaking TAs to better understand the different effects they impose on non-native English speaking students.

3.6 Tables & Figures

Table 3.1: Composition of TA languages

Language	Number of speakers	Number of students	Share in sample (%)
ITALIAN	8	1,427	16.53
MANDARIN	7	1,213	14.05
ENGLISH	4	965	11.18
GREEK	4	716	8.29
HINDI	4	678	7.85
WAD PAGGA	2	657	7.61
SPANISH	4	466	5.40
GERMAN	4	419	4.85
POLISH	1	350	4.05
PERSIAN	2	292	3.38
BULGARIAN	1	281	3.25
TURKISH	1	222	2.57
KOREAN	1	217	2.51
LITHUANIAN	1	200	2.32
PORTUGUESE	1	146	1.69
HEBREW	1	129	1.49
JAPANESE	1	96	1.11
FRENCH	1	91	1.05
SIAMESE	1	49	0.57
AZERBAIJANI	1	21	0.24
Total Sample	50	8,635	100

Notes: This table shows languages spoken by TAs, the number of speakers, number of students assigned to each language and share of students taught by the speaker of that language.

Table 3.2: TA sample descriptives

	Mean	*SD*	*Min*	*Max*
TA's age	28.77	3.80	22.00	41.00
TA's experience	1.50	0.74	1.00	4.00
No. of obs	*103*			
Female TA	0.47	0.50	0.00	1.00
TA's Ling. dist. to English	83.75	30.35	0.00	101.27
Italian-speaking TA	0.17	0.37	0.00	1.00
Mandarin-speaking TA	0.14	0.35	0.00	1.00
English-speaking TA	0.11	0.32	0.00	1.00
Other language TA	0.58	0.49	0.00	1.00
My own language TA	0.09	0.28	0.00	1.00
No. of obs				

Notes: This table summarises descriptive statistics of TAs and other associated variables of interest. To capture changing age and experience of the TAs, summary statistics of these variables are not for the whole sample, but rather for the total number of TAs between 2006 and 2016, this way counting one TA as two if they taught in two years of the sample.

Table 3.3: Testing for random assignment: do students choose TAs?

	Non-English TA	TA age	TA ld English	English TA	Mandarin TA	Italian TA	Female TA
Non-English speaker	0.022	-0.000	0.000	-0.022	-0.003	0.001	0.006
	(0.028)	(0.002)	(0.000)	(0.028)	(0.020)	(0.020)	(0.017)
first year GPA[a]	-0.616	-0.002	-0.006	0.616	0.545	-0.039	0.379
	(1.145)	(0.092)	(0.012)	(1.145)	(0.991)	(1.120)	(0.695)
Ling. dist. to English	2.211	-0.009	0.014	-2.211	-0.161	0.063	0.696
	(2.689)	(0.205)	(0.028)	(2.689)	(1.883)	(1.911)	(1.592)
Student's age	0.014	-0.001	0.000	-0.014	0.005	0.015	0.004
	(0.081)	(0.006)	(0.001)	(0.081)	(0.052)	(0.053)	(0.046)
Gender: Female	0.002	-0.000	0.000	-0.002	-0.005	0.002	-0.003
	(0.029)	(0.002)	(0.000)	(0.029)	(0.021)	(0.021)	(0.017)
Language: English	-0.022	0.000	-0.000	0.022	0.003	-0.001	-0.006
	(0.028)	(0.002)	(0.000)	(0.028)	(0.020)	(0.020)	(0.017)
Language: Mandarin	0.029	0.000	0.000	-0.029	0.010	0.008	0.012
	(0.021)	(0.002)	(0.000)	(0.021)	(0.015)	(0.015)	(0.013)
Language: Italian	0.004	0.000	0.000	-0.004	-0.003	0.003	-0.001
	(0.008)	(0.001)	(0.000)	(0.008)	(0.006)	(0.007)	(0.006)
share/mean (urn)	yes	yes	yes	yes	yes	yes	yes
Course $\times$ year FE	yes	yes	yes	yes	yes	yes	yes
Study program FE	yes	yes	yes	yes	yes	yes	yes
Day/Time FE	yes	yes	yes	yes	yes	yes	yes
No. of observations[a]	8635	8635	8635	8635	8635	8635	8635

Notes: This table summarises results of regressions of individual student characteristics on TA characteristics; each row and column represents a separate regression. Each regression includes the seminar-wise leave-me-out means/shares and the course/year-wise leave-me-out mean/share and a number of fixed effects. [a]First year GPA is only available for 4,404 observations and does not generally enter my later specifications. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at individual level, are reported in parentheses.

Table 3.4: Testing for random assignment at seminar level: do TAs choose students?

	Non-English TA	TA age	TA ld English	English TA	Mandarin TA	Italian TA	Female TA
Non-English speaker	-0.097	0.228	-13.329*	0.097	-0.161	-0.348*	0.233
	(0.082)	(1.421)	(7.974)	(0.082)	(0.129)	(0.182)	(0.184)
first year GPA[a]	-0.000	0.028*	-0.026	0.000	0.002	-0.001	0.000
	(0.001)	(0.017)	(0.109)	(0.001)	(0.002)	(0.001)	(0.002)
Ling. dist. to English	-0.000	0.004	-0.022	0.000	-0.000	0.000	0.000
	(0.000)	(0.004)	(0.018)	(0.000)	(0.000)	(0.000)	(0.000)
Student's age	0.008*	-0.049	1.021**	-0.008*	0.007	-0.014*	0.009
	(0.005)	(0.051)	(0.470)	(0.005)	(0.009)	(0.008)	(0.007)
Gender: Female	-0.034	0.977***	-3.392	0.034	0.005	0.059	-0.024
	(0.022)	(0.341)	(2.222)	(0.022)	(0.040)	(0.052)	(0.051)
Language: English	0.024	-0.414	2.163	-0.024	0.010	-0.032	-0.021
	(0.018)	(0.361)	(1.774)	(0.018)	(0.024)	(0.028)	(0.030)
Language: Mandarin	0.037	0.677	3.716	-0.037	-0.008	0.080*	0.035
	(0.030)	(0.437)	(2.884)	(0.030)	(0.053)	(0.048)	(0.064)
Language: Italian	-0.069	-0.209	-6.465	0.069	-0.102*	-0.085	0.233**
	(0.050)	(0.837)	(4.373)	(0.050)	(0.060)	(0.073)	(0.096)
share/mean (urn)	yes	yes	yes	yes	yes	yes	yes
Course × year FE	yes	yes	yes	yes	yes	yes	yes
Day/Time FE	yes	yes	yes	yes	yes	yes	yes
No. of observations[a]	333	333	333	333	333	333	333

Notes: This table summarises results of regressions of TA characteristics on seminar-wise leave-me-out means/shares of student characteristics; each row and column represents a separate regression. Each regression includes the course/year-wise leave-me-out mean/share and a number of fixed effects. [a]First year GPA is only available for 4,404 observations and does not generally enter my later specifications. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at TA level, are reported in parentheses.

Table 3.5: TA language distance to English

Sample	Grade	Fail	2:1
Total			
TA's Lg Distance to EN	-0.001	0.000**	-0.000
	(0.000)	(0.000)	(0.000)
Mean of dep. var.	0.000	0.166	0.401
R^2	0.48	0.24	0.35
No. of observations	8635	8635	8635
English-speakers			
TA's Lg Distance to EN	-0.000	0.000	-0.000
	(0.001)	(0.000)	(0.000)
Mean of dep. var.	0.079	0.145	0.426
R^2	0.45	0.22	0.35
No. of observations	3974	3974	3974
Non-English-speakers			
TA's Lg Distance to EN	-0.001	0.001***	0.000
	(0.001)	(0.000)	(0.000)
Mean of dep. var.	-0.067	0.185	0.380
R^2	0.51	0.28	0.38
No. of observations	4661	4661	4661
Course × year FE	yes	yes	yes
Study program FE	yes	yes	yes
Day/Time FE	yes	yes	yes
Seminar controls	yes	yes	yes
Individual controls	yes	yes	yes
TA controls	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables (standardised grade, indicator for failing a course, indicator for receiving an honour (60% or above)) on TAs' linguistic distance to English. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance to English and whether they are an English speaker or not as well as students' and their peers' GPA. Seminar controls are leave-me-out share of females, number of students and mean age. I also control for TAs' age, gender and experience. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at individual level, are reported in parentheses.

Table 3.6: Effects of different Language TAs

Sample	Grade	Fail	Honour
Total			
TA Italian Speaker	-0.067	0.050*	0.005
	(0.056)	(0.026)	(0.032)
TA Mandarin Speaker	-0.056	0.035	-0.017
	(0.057)	(0.026)	(0.032)
Other Language TA	-0.080*	0.051**	-0.010
	(0.047)	(0.021)	(0.026)
Mean of dep. var.	0.000	0.166	0.401
R^2	0.48	0.24	0.35
No. of observations	8635	8635	8635
English			
TA Italian Speaker	-0.001	0.015	0.006
	(0.083)	(0.037)	(0.048)
TA Mandarin Speaker	0.022	-0.015	-0.056
	(0.082)	(0.036)	(0.048)
Other Language TA	-0.032	0.010	-0.034
	(0.066)	(0.030)	(0.038)
Mean of dep. var.	0.079	0.145	0.426
R^2	0.45	0.22	0.35
No. of observations	3974	3974	3974
Non-English			
TA Italian Speaker	-0.129*	0.092***	-0.004
	(0.075)	(0.036)	(0.044)
TA Mandarin Speaker	-0.124	0.091**	0.015
	(0.080)	(0.038)	(0.044)
Other Language TA	-0.118*	0.088***	0.006
	(0.068)	(0.030)	(0.037)
Mean of dep. var.	-0.067	0.185	0.380
R^2	0.51	0.28	0.38
No. of observations	4661	4661	4661
Course × year FE	yes	yes	yes
Study program FE	yes	yes	yes
Day/Time FE	yes	yes	yes
Seminar controls	yes	yes	yes
Individual controls	yes	yes	yes
TA controls	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables (standardised grade, indicator for failing a course, indicator for receiving an honour (60% or above)) on language dummies of TAs; Italian, Mandarin and other languages. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance to English and whether they are an English speaker or not as well as students' and their peers' GPA. Seminar controls are leave-me-out share of females, number of students and mean age. I also control for TAs' age, gender and experience. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at individual level, are reported in parentheses.

Table 3.7: My own language TA

Sample	Grade	Fail	2:1
Total			
My own lang TA	-0.095***	0.028**	-0.033*
	(0.031)	(0.014)	(0.017)
Mean of dep. var.	0.000	0.166	0.401
R^2	0.48	0.24	0.35
No. of observations	8635	8635	8635
English			
My own lang TA	0.027	-0.008	0.035
	(0.065)	(0.030)	(0.038)
Mean of dep. var.	0.079	0.145	0.426
R^2	0.45	0.22	0.35
No. of observations	3974	3974	3974
Non-English			
My own lang TA	-0.041	0.026	-0.008
	(0.047)	(0.022)	(0.025)
Mean of dep. var.	-0.067	0.185	0.380
R^2	0.51	0.28	0.38
No. of observations	4661	4661	4661
Course × year FE	yes	yes	yes
Study program FE	yes	yes	yes
Day/Time FE	yes	yes	yes
Seminar controls	yes	yes	yes
Individual controls	yes	yes	yes
TA controls	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables (standardised grade, indicator for failing a course, indicator for receiving an honour (60% or above)) on a dummy indicating whether TA is speaking the same language as student. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance to English and whether they are an English speaker or not as well as students' and their peers' GPA. Seminar controls are leave-me-out share of females, number of students and mean age. I also control for TAs' age, gender and experience. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at individual level, are reported in parentheses.

Table 3.8: TA language distance to English on third year outcomes

	Share of Numerical courses	Difficulty	Grade	Fail	Honour
Total					
TA's Lg Distance to EN	-0.000	0.000	0.002***	-0.000**	0.001**
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Mean of dep. var.	0.536	0.051	0.000	0.036	0.621
R^2	0.38	0.63	0.41	0.06	0.29
No. of observations	7636	7636	7636	7636	7636
English					
TA's Lg Distance to EN	-0.000	0.000**	0.001	-0.000*	0.000
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
R^2	0.37	0.68	0.36	0.09	0.26
No. of observations	3603	3603	3603	3603	3603
Non-English					
TA's Lg Distance to EN	-0.000*	-0.000	0.003***	-0.000	0.001**
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
R^2	0.39	0.62	0.46	0.07	0.32
No. of observations	4033	4033	4033	4033	4033
Course × year FE	yes	yes	yes	yes	yes
Study program FE	yes	yes	yes	yes	yes
Day/Time FE	yes	yes	yes	yes	yes
Seminar leader FE	yes	yes	yes	yes	yes
Seminar controls	yes	yes	yes	yes	yes
Individual controls	yes	yes	yes	yes	yes
TA controls	yes	yes	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables regarding course choices in third year on TAs' linguistic distance to English. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance to English and whether they are an English speaker or not as well as students' and their peers' GPA. Seminar controls are leave-me-out share of females, number of students and mean age. I also control for TAs' age, gender and experience. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at individual level, are reported in parentheses.

Table 3.9: Effects of different language TAs on third year outcomes

	Share of Numerical courses	Difficulty	Grade	Fail	Honour
Total					
TA Italian Speaker	-0.015	-0.001	0.132*	-0.026	0.044
	(0.018)	(0.001)	(0.069)	(0.017)	(0.036)
TA Mandarin Speaker	-0.009	0.001	0.184***	-0.018	0.054*
	(0.017)	(0.002)	(0.062)	(0.015)	(0.033)
Other Language TA	-0.021	0.001	0.177***	-0.028**	0.056**
	(0.014)	(0.001)	(0.054)	(0.013)	(0.028)
Mean of dep. var.	0.536	0.051	0.000	0.036	0.621
R^2	0.38	0.63	0.41	0.06	
No. of observations	7636	7636	7636	7636	7636
English					
TA Italian Speaker	0.008	0.002	0.032	-0.022	-0.004
	(0.026)	(0.002)	(0.102)	(0.023)	(0.051)
TA Mandarin Speaker	0.005	0.003	0.099	-0.019	0.023
	(0.025)	(0.002)	(0.087)	(0.019)	(0.045)
Other Language TA	0.003	0.003**	0.059	-0.026	0.002
	(0.021)	(0.001)	(0.076)	(0.016)	(0.039)
Mean of dep. var.	0.488	0.050	0.144	0.032	0.695
R^2	0.37	0.68	0.36	0.09	0.26
No. of observations	3603	3603	3603	3603	3603
Non-English					
TA Italian Speaker	-0.036	-0.003	0.209**	-0.030	0.072
	(0.024)	(0.002)	(0.092)	(0.022)	(0.051)
TA Mandarin Speaker	-0.032	-0.001	0.276***	-0.016	0.083*
	(0.023)	(0.002)	(0.086)	(0.022)	(0.048)
Other Language TA	-0.043**	-0.002	0.289***	-0.028	0.101**
	(0.020)	(0.002)	(0.075)	(0.019)	(0.042)
Mean of dep. var.	0.579	0.052	-0.129	0.040	0.555
R^2	0.39	0.62	0.46	0.07	0.32
No. of observations	4033	4033	4033	4033	4033
Course $\times$ year FE	yes	yes	yes	yes	yes
Study program FE	yes	yes	yes	yes	yes
Day/Time FE	yes	yes	yes	yes	yes
Seminar leader FE	yes	yes	yes	yes	yes
Seminar controls	yes	yes	yes	yes	yes
Individual controls	yes	yes	yes	yes	yes
TA controls	yes	yes	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables regarding course choices in third year on language dummies of TAs; Italian, Mandarin and other languages. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance to English and whether they are an English speaker or not as well as students' and their peers' GPA. Seminar controls are leave-me-out share of females, number of students and mean age. I also control for TAs' age, gender and experience. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at individual level, are reported in parentheses.

Table 3.10: My own language TA effects on third year outcomes

	Share of Numerical courses	Difficulty	Grade	Fail	Honour
Total					
My own lang TA	0.011	-0.000	-0.049	-0.004	-0.020
	(0.009)	(0.001)	(0.035)	(0.008)	(0.018)
Mean of dep. var.	0.536	0.051	0.000	0.036	0.621
R^2	0.38	0.63	0.41	0.06	0.29
No. of observations	7636	7636	7636	7636	7636
English					
My own lang TA	-0.003	-0.003**	-0.064	0.025	-0.004
	(0.021)	(0.001)	(0.076)	(0.016)	(0.039)
R^2	0.37	0.68	0.36	0.09	0.26
No. of observations	3603	3603	3603	3603	3603
Non-English					
My own lang TA	0.035**	0.002	-0.044	-0.015	-0.044
	(0.016)	(0.001)	(0.056)	(0.015)	(0.029)
R^2	0.38	0.62	0.45	0.07	0.31
No. of observations	4033	4033	4033	4033	4033
Course × year FE	yes	yes	yes	yes	yes
Study program FE	yes	yes	yes	yes	yes
Day/Time FE	yes	yes	yes	yes	yes
Seminar leader FE	yes	yes	yes	yes	yes
Seminar controls	yes	yes	yes	yes	yes
Individual controls	yes	yes	yes	yes	yes
TA controls	yes	yes	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables regarding course choices in third year on a dummy indicating whether TA is speaking the same language as student. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance to English and whether they are an English speaker or not as well as students' and their peers' GPA. Seminar controls are leave-me-out share of females, number of students and mean age. I also control for TAs' age, gender and experience. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at individual level, are reported in parentheses.

Table 3.11: TA Gender

Sample	Grade	Fail	2:1
Total			
Female TA	-0.034	0.004	-0.014
	(0.027)	(0.012)	(0.015)
Mean of dep. var.	0.000	0.166	0.401
R^2	0.47	0.24	0.35
No. of observations	8635	8635	8635
Males			
Female TA	-0.089**	0.012	-0.041**
	(0.037)	(0.016)	(0.019)
Mean of dep. var.	-0.042	0.184	0.386
R^2	0.49	0.26	0.35
No. of observations	5088	5088	5088
Females			
Female TA	0.072*	-0.018	0.038*
	(0.042)	(0.019)	(0.023)
Mean of dep. var.	0.060	0.141	0.424
R^2	0.47	0.24	0.38
No. of observations	3547	3547	3547
Course × year FE	yes	yes	yes
Study program FE	yes	yes	yes
Day/Time FE	yes	yes	yes
Seminar controls	yes	yes	yes
Individual controls	yes	yes	yes
TA controls	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables (standardised grade, indicator for failing a course, indicator for receiving an honour (60% or above)) on a dummy indicating TAs' gender. Results by students' gender are derived from split sample models. Individual controls contain age, linguistic distance to English and whether they are an English speaker or not as well as students' and their peers' GPA. Seminar controls are leave-me-out share of females, number of students and mean age. I also control for TAs' age, linguistic distance to English and experience. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at individual level, are reported in parentheses.

Table 3.12: TA gender effects on third year outcomes

	Share of Numerical courses	Difficulty	Grade	Fail	Honour
Total					
TA female	-0.001	-0.000	-0.013	-0.008	-0.008
	(0.008)	(0.001)	(0.031)	(0.006)	(0.016)
Mean of dep. var.	0.536	0.051	0.000	0.036	0.621
R^2	0.38	0.63	0.41	0.06	0.29
No. of observations	7636	7636	7636	7636	7636
Males					
TA female	-0.000	0.000	0.014	-0.013	0.002
	(0.010)	(0.001)	(0.042)	(0.008)	(0.021)
R^2	0.38	0.66	0.43	0.07	0.31
No. of observations	4466	4466	4466	4466	4466
Females					
TA female	-0.001	-0.000	-0.037	-0.001	-0.013
	(0.012)	(0.001)	(0.044)	(0.011)	(0.026)
R^2	0.43	0.62	0.42	0.08	0.29
No. of observations	3170	3170	3170	3170	3170
Course $\times$ year FE	yes	yes	yes	yes	yes
Study program FE	yes	yes	yes	yes	yes
Day/Time FE	yes	yes	yes	yes	yes
Seminar leader FE	yes	yes	yes	yes	yes
Seminar controls	yes	yes	yes	yes	yes
Individual controls	yes	yes	yes	yes	yes
TA controls	yes	yes	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables regarding course choices in third year on a dummy indicating TAs' gender. Results by students' gender are derived from split sample models. Individual controls contain age, linguistic distance to English and whether they are an English speaker or not as well as students' and their peers' GPA. Seminar controls are leave-me-out share of females, number of students and mean age. I also control for TAs' age, linguistic distance to English and experience. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at individual level, are reported in parentheses.

Table 3.13: Robustness: Effects of different Language TAs without ability controls

Sample	Grade	Fail	Honour
Total			
TA Italian Speaker	-0.069	0.051*	0.005
	(0.074)	(0.028)	(0.037)
TA Mandarin Speaker	0.006	0.023	0.008
	(0.076)	(0.028)	(0.038)
Other Language TA	-0.049	0.045*	0.002
	(0.064)	(0.023)	(0.031)
Mean of dep. var.	0.000	0.166	0.401
R^2	0.05	0.07	0.12
No. of observations	8635	8635	8635
English			
TA Italian Speaker	-0.022	0.020	-0.002
	(0.104)	(0.040)	(0.055)
TA Mandarin Speaker	0.070	-0.026	-0.035
	(0.103)	(0.038)	(0.055)
Other Language TA	-0.011	0.005	-0.024
	(0.084)	(0.032)	(0.043)
Mean of dep. var.	0.079	0.145	0.426
R^2	0.07	0.09	0.14
No. of observations	3974	3974	3974
Non-English			
TA Italian Speaker	-0.125	0.092**	-0.003
	(0.106)	(0.040)	(0.052)
TA Mandarin Speaker	-0.037	0.073*	0.046
	(0.114)	(0.042)	(0.052)
Other Language TA	-0.071	0.079**	0.023
	(0.097)	(0.034)	(0.045)
Mean of dep. var.	-0.067	0.185	0.380
R^2	0.08	0.09	0.15
No. of observations	4661	4661	4661
Course × year FE	yes	yes	yes
Study program FE	yes	yes	yes
Day/Time FE	yes	yes	yes
Seminar controls	yes	yes	yes
Individual controls	yes	yes	yes
TA controls	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables (standardised grade, indicator for failing a course, indicator for receiving an honour (60% or above)) on language dummies of TAs; Italian, Mandarin and other languages. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance to English and whether they are an English speaker or not. Seminar controls are leave-me-out share of females, number of students and mean age. I also control for TAs' age, gender and experience. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at individual level, are reported in parentheses.

Table 3.14: Robustness: Different language TAs on third year outcomes without ability controls

	Share of Numerical courses	Difficulty	Grade	Fail	Honour
Total					
TA Italian Speaker	-0.014	-0.001	0.145*	-0.027	0.047
	(0.019)	(0.001)	(0.085)	(0.017)	(0.040)
TA Mandarin Speaker	-0.004	0.001	0.237***	-0.022	0.068*
	(0.018)	(0.001)	(0.079)	(0.015)	(0.037)
Other Language TA	-0.017	0.001	0.219***	-0.030**	0.069**
	(0.015)	(0.001)	(0.069)	(0.013)	(0.032)
Mean of dep. var.	0.536	0.051	0.000	0.036	0.621
R^2	0.32	0.62	0.10	0.04	
No. of observations	7636	7636	7636	7636	7636
English					
TA Italian Speaker	-0.002	0.001	-0.049	-0.018	-0.032
	(0.027)	(0.002)	(0.119)	(0.023)	(0.055)
TA Mandarin Speaker	-0.004	0.002	0.053	-0.018	0.001
	(0.026)	(0.002)	(0.106)	(0.019)	(0.049)
Other Language TA	-0.004	0.003**	0.012	-0.024	-0.017
	(0.022)	(0.001)	(0.090)	(0.017)	(0.041)
Mean of dep. var.	0.488	0.050	0.144	0.032	0.695
R^2	0.32	0.67	0.10	0.07	0.13
No. of observations	3603	3603	3603	3603	3603
Non-English					
TA Italian Speaker	-0.025	-0.003	0.305***	-0.035	0.109*
	(0.025)	(0.002)	(0.118)	(0.022)	(0.059)
TA Mandarin Speaker	-0.014	-0.001	0.440***	-0.024	0.145***
	(0.024)	(0.002)	(0.114)	(0.022)	(0.055)
Other Language TA	-0.027	-0.002	0.423***	-0.035*	0.152***
	(0.022)	(0.002)	(0.102)	(0.019)	(0.049)
Mean of dep. var.	0.579	0.052	-0.129	0.040	0.555
R^2	0.32	0.61	0.12	0.05	0.11
No. of observations	4033	4033	4033	4033	4033
Course × year FE	yes	yes	yes	yes	yes
Study program FE	yes	yes	yes	yes	yes
Day/Time FE	yes	yes	yes	yes	yes
Seminar leader FE	yes	yes	yes	yes	yes
Seminar controls	yes	yes	yes	yes	yes
Individual controls	yes	yes	yes	yes	yes
TA controls	yes	yes	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables regarding course choices in third year on language dummies of TAs; Italian, Mandarin and other languages. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance to English and whether they are an English speaker or not. Seminar controls are leave-me-out share of females, number of students and mean age. I also control for TAs' age, gender and experience. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Bibliography

Abadie, A., Athey, S., Imbens, G. and Wooldridge, J. (2017). When Should You Adjust Standard Errors for Clustering? *NBER Working Paper* 24003.

Abraham, K. and Katz, L. (1986). Cyclical unemployment: Sectoral shifts or aggregate disturbances? *Journal of Political Economy* 94: 507–522.

Ager, P. and Brückner, M. (2013). Cultural diversity and economic growth: Evidence from the us during the age of mass migration. *European Economic Review* 64: 76–97.

Alesina, A. and Ferrara, E. L. (2005). Ethnic Diversity and Economic Performance. *Journal of Economic Literature* 43: 762–800.

Alesina, A., Harnoss, J. and Rapoport, H. (2016). Birthplace diversity and economic prosperity. *Journal of Economic Growth* 21: 101–138.

Alesina, A. and La Ferrara, E. (2002). Who trusts others? *Journal of Public Economics* 85: 207–234.

Algan, Y., Hemet, C. and Laitin, D. D. (2016). The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation. *Journal of Political Economy* 124: 696–733.

Altonji, J. G., Elder, T. E. and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy* 113(1): 151–184.

Anelli, M. and Peri, G. (2019). The Effects of High School Peers' Gender on College Major, College Performance and Income. *Economic Journal* 129: 553–602.

Anelli, M., Shih, K. and Williams, K. (2017). Foreign Peer Effects and STEM Major Choice. *IZA Discussion Paper Series* 10743.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics* 30: 98–108.

Baker, M. (1992). Unemeployment duration: Composition effects and cyclical variability. *American Economic Review* 82: 313–321.

Ballatore, R. M., Fort, M. and Ichino, A. (2018). Tower of Babel in the Classroom: Immigrants and Natives in Italian Schools. *Journal of Labor Economics* 36.

Barnichon, R. and Figura, A. (2015). Labor market heterogeneity and the aggregate matching function. *American Economic Journal: Macroeconomics* 7: 222–249.

Betts, J. R. and Fairlie, R. W. (2003). Does immigration induce 'native flight' from public schools into private schools? *Journal of Public Economics* 87: 987–1012.

Blanchard, O. and Diamond, P. (1989). The beveridge curve. *Brookings Papers on Economic Activity* 1989: 1–76.

Blundell, R., Crawford, C. and Jin, W. (2014). What can wages and employment tell us about the uk's productivity puzzle. *The Economic Journal* 124: 377–407.

Boisjoly, J., Duncan, G., Kremer, M., Levy, D. and Eccles, J. (2006). Empathy or Antipathy? The Impact of Diversity. *American Economic Review* 96: 1890–1905.

Borjas, G. J. (2000). Foreign-born teaching assistants and the academic performance of undergraduates. *American Economic Review* 90: 355–359.

Borowczyk-Martins, D., Jolivet, G. and Postel-Vinay, F. (2013). Accounting for endogeneity in matching function estimation. *Review of Economic Dynamics* 16: 440–451.

Braakmann, N. and McDonald, S. (2018). The Impact of Student Diversity on Student Outcomes at University and Beyond - Evidence From English Administrative Data. *Unpublished manuscript* .

Bredtmann, J. and Smith, N. (2018). Inequalities in educational outcomes: How important is the family? *Oxford Bulletin of Economics and Statistics* 80: 1117—1144.

Britton, J., Belfield, C., Erve, L. van der, Vignoles, A., Walker, I. and Dearden, L. (2018). Estimating the Returns to College Degrees Using Linked Administrative Data. *Institute for Fiscal Studies Research Report* .

Brunello, G. and Rocco, L. (2013). The Effect of Immigration on the School Performance of Natives: Cross Country Evidence Using PISA Test Scores. *Economics of Education Review* 32: 234–246.

Caeyers, B. and Fafchamps, M. (2016). Exclusion Bias in the Estimation of Peer Effects. NBER Working Papers 22565, National Bureau of Economic Research, Inc.

Carrell, S. E., Hoekstra, M. and West, J. E. (forthcoming). The impact of college diversity on behavior toward minorities. *American Economic Journal: Economic Policy* .

Carrell, S. E., Page, M. E. and West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics* 125: 1101–1144.

Dale-Olsen, H. and Finseraas, H. (2019). Linguistic diversity and workplace productivity .

Davis, S., Faberman, J. and Haltiwanger, J. (2013). The establishment-level behavior of vacancies and hiring. *The Quarterly Journal of Economics* 128: 581–622.

Diamond, P. (1982). Wage determination and inefficiency in search equilibrium. *Review of Economic Studies* 49: 217–227.

Diette, T. M. and Oyelere, R. U. (2017). Do limited english students jeopardize the education of other students? lessons from the north carolina public school system. *Education Economics* 25:5: 446–461.

Dustmann, C. and Glitz, A. (2011). Migration and education 4: 327–439.

Elsby, M., Hobijn, B. and Sahin, A. (2010). The labour market in the great recession. *Brookings Papers on Economic Activity* .

Evans, M. O. (1992). An estimate of race and gender role-model effects in teaching high school. *The Journal of Economic Education* 23: 209–217.

Faberman, R. J. and Kudlyak, M. (2019). The intensity of job search and search duration. *American Economic Journal: Macroeconomics* 11: 327–357.

Fairlie, R. W., Hoffmann, F. and Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *The American Economic Review* 104: 2567–2591.

Feld, J. and Zölitz, U. (2017). Understanding Peer Effects: On the Nature, Estimation and Channels of Peer Effects. *Journal of Labor Economics* 35: 387–428.

Felda, J., Salamancab, N. and Zölitzc, U. (2019). Students are almost as effective as professors in university teaching. *CESifo Working Paper No. 7764* .

Figlio, D. N. and Özek, U. (2017). Unwelcome Guests? The Effects of Refugees on the Educational Outcomes of Incumbent Students. NBER Working Papers 23661, National Bureau of Economic Research, Inc.

Fleisher, B., Hashimoto, M. and Weinberg, B. A. (2002). Foreign gtas can be effective teachers of economics. *The Journal of Economic Education* 33: 299–325.

Frattini, T. and Meschi, E. (2017). The Effect of Immigrant Peers in Vocational Schools. *IZA Discussion Paper Series* 11027.

Gavazza, A., Mongey, S. and Violante, G. L. (2014). What shifts the beveridge curve? recruitment effort and financial shocks. *2014 Meeting Papers from Society for Economic Dynamics* .

Geay, C., McNally, S. and Telhaj, S. (2013). Non-Native Speakers of English in the Classroom: What Are the Effects on Pupil Performance? *The Economic Journal* 123: 281–307.

Gould, E. D., Lavy, V. and Paserman, M. D. (2009). Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi Experimental Evidence. *The Economic Journal* 119: 1243–1269.

Guryan, J., Kroft, K. and Notowidigdo, M. J. (2009). Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments. *American Economic Journal: Applied Economics* 1: 34–68.

Hall, R. E. and Schulhofer-Wohl, S. (2018). Measuring job-finding rates and matching efficiency with heterogeneous job-seekers. *American Economic Journal: Macroeconomics* 10: 1–32.

Hobijn, B. and Sahin, A. (2013). Beveridge curve shifts across countries since the great recession. *IMF Economic Review* 61: 566–600.

Hoffmann, F. and Oreopoulos, P. (2009). A professor like me: The influence of instructor gender on college achievement. *The Journal of Human Resources* 44: 479–494.

Hoogendoorn, S. and Van Praag, M. (2012). Ethnic diversity and team performance: A field experiment. Tech. rep., inbergen Institute Discussion Paper No. 12-068/3.

Hunt, J. (forthcoming). The impact of immigration on the educational attainment of natives. *Journal of Human Ressources* .

Isphording, I. E. and Otten, S. (2014). Linguistic Barriers in the Destination Language Acquisition of Immigrants. *Journal of Economic Behavior and Organization* 105: 30–50.

Jensen, P. and Rasmussen, A. W. (2011). The Effect of Immigrant Concentration in Schools on Native and Immigrant Children's Reading and Math Skills. *Economics of Education Review* 30: 1503–1515.

Kahane, L., Longley, N. and Simmons, R. (2013). The Effects of Coworker Heterogeneity on Firm-Level Output: Assessing the Impacts of Cultural and Language Diversity in the National Hockey League. *The Review of Economics and Statistics* 95: 302–314.

Kroft, K., Lange, F., Notowidigdo, M. J. and Katz, L. F. (2016). Long-term unemployment and the great recession: the role of composition, duration dependence, and non-participation. *Journal of Labor Economics* 34: 7–54.

Lazear, E. P. (1999). Culture and Language. *Journal of Political Economy* 107: S95–S126.

Lusher, L., Campbellb, D. and Carrel, S. (2018). Tas like me: Racial interactions between graduate teaching assistants and undergraduates. *Journal of Public Economics* 159: 203–224.

Machin, S. and Murphy, R. (2018). Paying Out and Crowding Out? The Globalisation of Higher Education. *Journal of Economic Geography. forthcoming* .

Maestri, V. (2017). Can Ethnic Diversity Have a Positive Effect on School Achievement? *Education Economics* 25: 290–303.

McHenry, P. (2015). Immigration and the Human Capital of Natives. *Journal of Human Resources* 50: 34–71.

Migration Advisory Committee (2018). Impact of international students in the UK. Tech. rep.

Mortensen, D. and Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. *Review of Economic Studies* 61: 397–415.

Ohinata, A. and Ours, J. C. van (2013). How Immigrant Children Affect the Academic Achievement of Native Dutch Children. *Economic Journal* 123: 308–331.

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* .

Ottaviano, G. I. and Peri, G. (2006). The economic value of cultural diversity: evidence from us cities. *Journal of Economic geography* 6: 9–44.

Petrongolo, B. and Pissarides, C. A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature* 39: 390–431.

Pianta, R. C. (1994). Patterns of relationships between children and kindergarten teachers. *Journal of School Psychology* 32: 15–31.

Pissarides, C. A. (1985). Short-run equilibrium dynamics of unemployment, vacancies, and real wages. *The American Economic Review* 75: 676–690.

Pissarides, C. A. (1986). Unemployment and vacancies in britain. *Economic Policy* 1: 499–559.

Pissarides, C. A. (2000). *Equilibrium Unemployment Theory*. MIT Press, 2nd ed.

Pizzinelli, C. and Speigner, B. (2017). Matching efficiency and labour market heterogeneity in the united kingdom. *BoE Working Paper* .

Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twenty-first century the 2006 johan skytte prize lecture. *Scandinavian political studies* 30: 137–174.

Sahin, A., Patterson, C., Topa, G. and Violante, G. L. (2013). Mismatch unemployment in the uk. *Working Paper* .

Sahin, A., Song, J., Topa, G. and Violante, G. L. (2014). Mismatch unemployment. *American Economic Review* 104: 3529–3564.

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review* 95: 25–49.

Shimer, R. (2008). The probability of finding a job. *American Economic Review* 98: 268–273.

Smith, C. J. (2012). Unemployment and mismatch in the uk. *Working Paper* .

Top-Universities, Q. (2018). World university ranking 2018 - economics / econometrics.

Trax, M., Brunow, S. and Suedekum, J. (2015). Cultural Diversity and Plant-level Productivity. *Regional Science and Urban Economics* 53.

UNESCO (2018). Outbound internationally mobile students by host region.

Veracierto, M. (2011). Worker flows and matching efficiency. *Economic Perspectives* 35: 147–169.

White, K. M. (2013). Associations between teacher–child relationships and children's writing in kindergarten and first grade. *Early Childhood Research Quarterly* 28: 166–176.

Wichmann, S., Holman, E. W. and Brown, C. H. (2018). The ASJP Database. *version 18* .

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics* 141: 1281–1301.

Appendix A

Chapter 1 Appendix: Administrative Data Analysis and Comparison

A.1 Data description

A.1.1 Unemployment

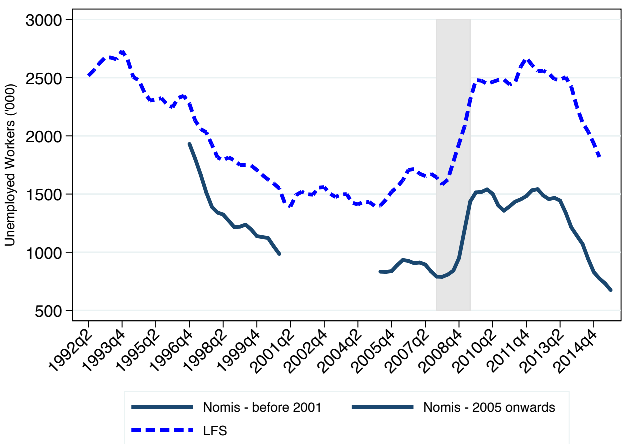


Figure A.1: Aggregate Unemployment (Nomis and LFS, Levels)

Note: Nomis Claimant Counts here is a quarterly average of monthly data. Both Nomis and LFS unemployment is smoothed over 2 quarters. Shaded areas represent recessions indicated by FRED (OECD).

Alternative source of unemployment data in the UK is Job Seeker's Allowance Claimant Count (Nomis). LFS unemployment and the claimant counts are consistent to a high degree. However, LFS unemployment measure is generally higher and more representative of the true population in the UK. Figure A.1 graphs both unemployment series. Correlation between the two measures is 0.9846 before the break in the data in October 2000 and it is 0.9462 after January 2005. Both unemployment series overlap to some extent: claimants are generally recorded as unemployed in International Labour Organization (ILO) definition of unemployment (see Table A.1 for details). However, non-claimants can appear among unemployed if they are not eligible for benefits for one of the below reasons[1]:

- They are only looking for part-time work;
- They are under 18 and are looking for work but do not take up the offer of a Youth Training place;
- They are students looking for vacation work;
- They have left their job voluntarily.

Analogously, some people recorded in the claimant count would not be counted as unemployed. People can claim Jobseeker's Allowance if they earn low income from part-time work and therefore these people would not be unemployed. Table A.1 summarises the main features of the two data sources.

Alternative measures of unemployment

The International Labour Organization (ILO) definition of unemployment is limited in some sense because it ignores those who are out of the labour force. The alternative measures suggested by the Bureau of Labor Statistics (BLS) allows to deepen the understanding of true unemployment situation in the UK. In this paper, all six unemployment rates are produced using the LFS data. U1-U2 give narrower definition of unemployment, U3 is the official ILO definition, while U4-U6 give broader concept of unemployment[2]. U1 gives the number of long-term (15 weeks or longer) unemployed workers as a percent of the labour force. The closest alternative in the LFS is unemployment of 3 months or longer. The broader definitions, U4-U6, include discouraged or marginally attached workers. Discouraged workers fall as a part of marginally attached. Discouraged are those workers who are not in the labour force, they would like to and are able to work, they have looked for work in the past 12 months, but not in the past 4 weeks, because they believe that there are no jobs available for them. Marginally attached are all those able and willing to work that were not looking for work in the past 4 weeks for any reason. All broader measures represent potential groups of workers who under certain conditions would work and therefore in some ways they can be considered as unemployed.

[1] https://www.detini.gov.uk/sites/default/files/publications/deti/Summary%20LFS%20CC%20explanation%20for%20the%20web.pdf (accessed on 9 January 2016)

[2] Details and definitions of BLS alternative measures of unemployment can be found here: http://www.bls.gov/lau/stalt.htm (accessed on 9 January 2016).

Table A.1: Specification of Unemployment Data in the UK

	JSA Claimant Count	LFS
Type	Administrative	Household Survey
Definition of unemployment	Jobseekers that are out of work, capable of, available for and actively seeking work during the week in which their claim is made	ILO - Answer 'yes' to both 'whether the respondent is available to work in the next 2 weeks' and 'whether he/she has looked for work in the last 4 weeks'
Period	From July 1996 and regularly updated. Break between October 2000 and January 2005. Quarterly before October 2000, monthly from January 2005	Quarterly from 1992 and updated regularly
Sample	All JSA claimants	50000 households every quarter
Worker characteristics	age gender claim duration	age gender unemployment duration marital status ethnicity qualifications
Labour market characteristics	region (UK excl. NI) occupation (1-digit SOC2000)	region (UK excl. NI) occupation (1-digit SOC1990) industry (1-digit SIC2007)
Collection period	Second Thursday of a given month	Respondents are interviewed over 13 weeks in a given quarter and are asked about their situation and activities in a reference week (a seven day period that ends on a Sunday). Most of the interviews are carried out in the week following the reference week

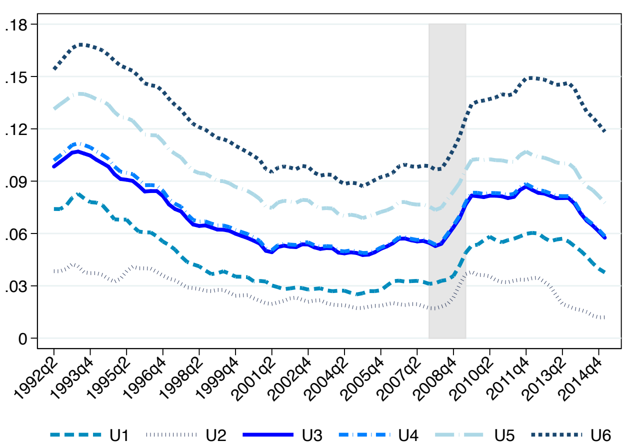


Figure A.2: Alternative Measures of Unemployment

Note: All measures are constructed using LFS. 2-quarter moving averages. These measures were constructed using Bureau of Labor Statistics definitions. Shaded area represents the Great Recession as indicated by FRED (OECD).

Figure A.2 plots U1-U6 unemployment rates. All six alternative measures generally move together. There was a sharp increase in all unemployment rates during the Great Recession. U6 - the broadest unemployment measure, which also includes part-time workers who are not working full-time for economic reasons - surged from around 10% in the beginning of the recession to as high as 15% in the aftermath of the recession.

A.1.2 Matches

Nomis - Claimant Off-flows and Vacancy Outflows

Nomis provides two plausible measures of total matches. *Claimant off-flows*, the number of people who stop claiming Jobseeker's Allowance, is one of them. However, it is not always true that unemployed workers stop claiming benefits because they found a job. They might do so for other reasons, such as claiming benefits for a maximum period of six months, or a change in other circumstances that make claimants ineligible for JSA. Therefore, claimant off-flows, as a measure of total matches, is subject to measurement error.

Another measure of new hires in the labour market is *vacancy outflow*, which is the count of vacancies that have either been filled by JobCentre Plus or withdrawn by employers. Similarly as with claimant off-flows, we cannot assume that all vacancies were filled by unemployed workers. Some of the jobs might have been taken by people out of the labour force, or workers who experienced job-to-job transitions (without facing unemployment).

It is not possible to correct for these measurement errors, however, as suggested by Sahin et al.(2013) in their working paper, one might want to take the average of the two possible measures for the estimation of the matching function.

Data specifications, that apply to JSA claimant count and JCP vacancies (i.e., type, period, collection period, worker and labour market characteristics given in Tables A.1 and A.2) are also true for claimant off-flows and vacancy outflows.

Figure A.3 plots new hires (a) and job finding rate (b) using both claimant off-flows and vacancy outflows. Interestingly, the correlation between the two new matches series is -0.4520. Midway through the Great Recession, the number of people who stopped claiming JSA benefits started to increase. With a sharp increase in the number of claimants, the count of people who run out of benefits (claim for the maximum period) is also going up. This may explain an increase in the number of matches given by claimant off-flows.

Panel (b) in Figure A.3 shows the monthly job finding rates constructed using claimant off-flows and vacancy outflows. Both rates generally move together, the correlation between the two is 0.8715. Before and during the Great Recession, however, vacancy outflows give much higher job finding probability. Both series

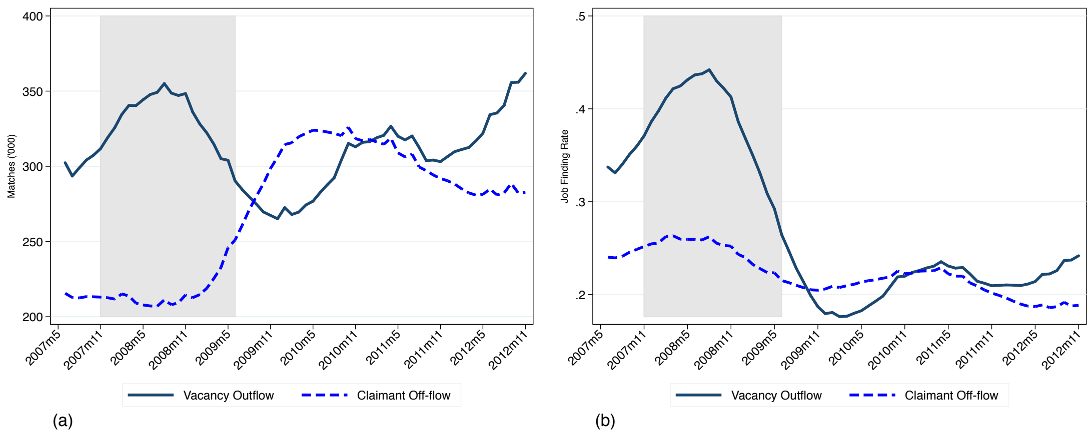


Figure A.3: Matches and Job Finding Rate (Nomis)

Note: All series are 12-month moving averages. Shaded areas represent recessions indicated by FRED (OECD).

show a decline in the job finding rate during the recession.

A.2 Alternative LFS measure of the job finding rate

In addition to UE transitions as a measure of the number of matches between vacant jobs and unemployed workers, I construct an alternative measure. If a person is employed for 3 or less months, I count this as a new match. However, I need to account for job-to-job transitions. I observe people who had jobs in the previous quarter and who again are employed in the current quarter. If they left a paid job in the last 3 months, this is counted as a job-to-job transition. This measure, however, does not account for inactivity to employment transitions as it is not possible to tell where the newly employed people are coming from and if they were unemployed last term. Therefore, I argue that UE transitions is a better measure of the new matches in the labour market.

Figure A.4 plots both LFS matches series (left) and job finding rates (right).

The UE transitions measure of new hires is generally below the short tenure measure of matches. Over the whole period, the correlation between the two series is 0.7065. The number of new hires constructed from the short tenure matches being consistently above the UE transitions measure can be explained as follows. In the UE case, if a person is recorded as employed last quarter and is again employed this quarter, there is no new job finding recorded. However, during the 3-month period, the same person might have experienced a short unemployment spell, and in the short tenure data analysis, this would be captured and counted as a new match.

The graph on the right side of Figure A.4 shows job finding rates using both LFS measures of matches. Both series move together and record a sharp decrease in the probability of finding a new job during and after the Great Recession. The

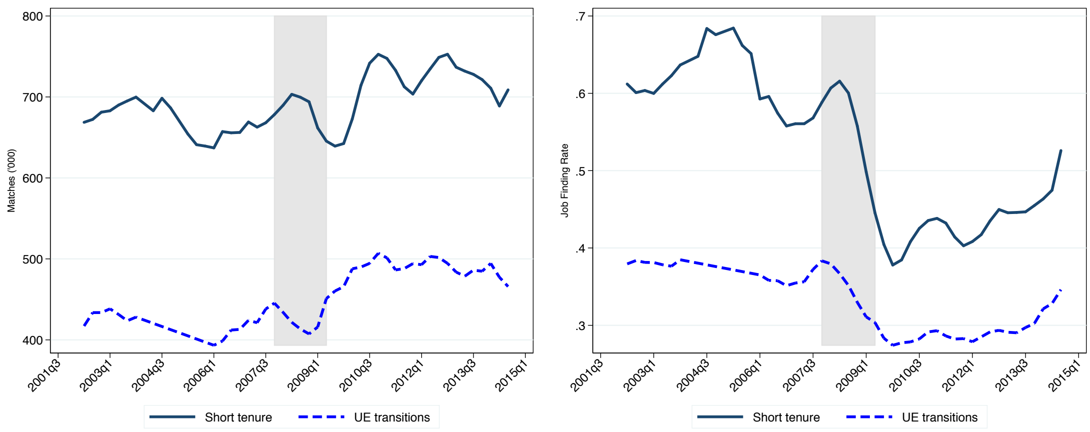


Figure A.4: Matches and Job Finding Rate (LFS)

Note: All series are 4-quarter moving averages. Shaded area represents the Great Recession as indicated by FRED (OECD).

correlation coefficient between the two rates is 0.9794.

A.2.1 Vacancies

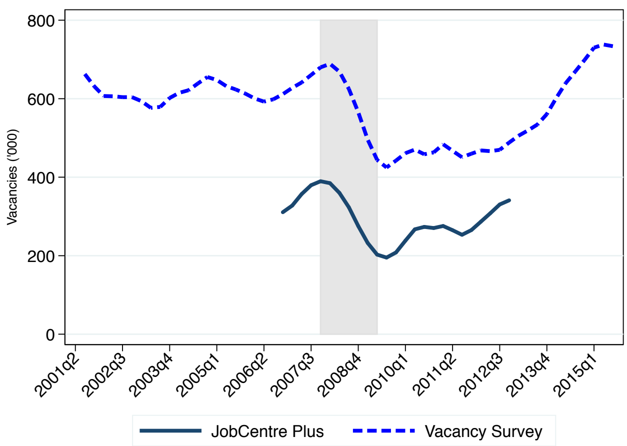


Figure A.5: Aggregate Vacancies (Nomis and Vacancy Survey, Levels)

Note: Nomis JobCentre Plus vacancies and Vacancy Survey here are quarterly averages of monthly data. Both Nomis and Vacancy Survey series are deseasonalised and smoothed over 2 quarters. Shaded areas represent recessions indicated by FRED (OECD).

Similarly as with unemployment data, vacancy series are also available from two different data sources. In addition to Vacancy Survey, there is also Administrative JobCentre Plus (JCP) vacancy statistics, which come from Nomis. Table A.2 provides a summary of vacancy data specifics in the UK.

JCP is the Public Employment Service for Great Britain that accounts for only about one third of the vacancies in the UK. The rest is advertised by employment

Table A.2: Specification of Vacancy Data in the UK

	JobCentre Plus	**Vacancy Survey**
Type	Administrative - supplied by Department for Work and Pensions (DWP)	Business Survey
Period	From April 1994 until November 2012. Break between October 2000 and April 2004. Quarterly before October 2000, monthly from April 2004	Monthly from April 2001 and updated regularly
Sample	All vacancies notified to Job Centres	6000 businesses every month, population of 1.93 million
Labour market characteristics	region (UK excl. NI) occupation (1-digit SOC2000) industry (1-digit SIC2003)	industry (1-digit SIC2007, excl. Agriculture, Forestry and Fishing) business size
Collection period	First Friday of a given month	Data is collected over 16 working days starting the first Friday of the month, unless it is the first day of the month, then the reference day is moved to the second Friday of the month

agencies or directly through employers. JobCentre Plus is a nation-wide employment support service and so it is very plausible that the jobs advertised through this service are targeted at the lower end of the professions scale in terms of skill requirements. Similarly as with claimant count data, JCP vacancies were also discontinued for a period of time, however, due to the measurement differences, the data before and after the break are not compatible. In addition, the procedures for recording and handling vacancies were changed in May 2006. From this point in time, a date of vacancy closure is agreed with the employer at the time of vacancy notification and therefore jobs are automatically withdrawn unless the employer advises that a later closure date is required. Over time, this would reduce the number of unfilled vacancies. To avoid measurement error coming from these changes, the analysis of JCP data starts in July 2006.

Vacancy Survey is more representative of a real vacancy creation situation in the UK. It is not only because of a broader occupational range; in fact, no employer is obligated to notify their vacancies to Job Centres and therefore JCP measure of vacancies is generally below Vacancy Survey. Figure A.5 shows plots of the two vacancy series. After the change in JPC vacancy handling, the correlation between the two measures is 0.8284.

A.3 Alternative data analysis

Figure A.6 shows the plots of the predicted job finding rates using the administrative Nomis data. The results presented in (a) and (c) graphs are from the matching function estimation using claimant off-flows as a measure of total matches. The standard aggregate matching function estimated over the whole period (July 2006 - November 2011) explains the movements of the job finding rate very well before and during the Great Recession. However, it underestimates the job finding probability in the aftermath of the recession until the next recession in the end of 2011, when the result is the opposite and the matching function over-predicts the job finding rate. This also holds for the estimation of the matching function before the peak of the Great Recession - the collapse of Lehman Brothers in September 2008 (c).

The bottom two graphs ((b) and (d)) repeat the estimation of the matching function using vacancy outflows as a measure of matches. In this case, the predicted job finding rate turns out to be very close to the actual data. However, the matching function estimated prior to September 2008 under-predicts the job finding probability after the Great Recession.

Table A.3 provides the estimated coefficients of the aggregate matching function. The elasticities from the estimations using claimant off-flows and vacancy outflows as measures of total matches are very different. Claimant off-flows give a more consistent estimate of the elasticity with respect to vacancies to the one found by Pissarides (1986). It is 0.214 for the whole period estimation compared to 0.3. When vacancy outflows are used as a measure of total matches, the estimated elasticity is 0.709 (regression (2) in Table A.3).

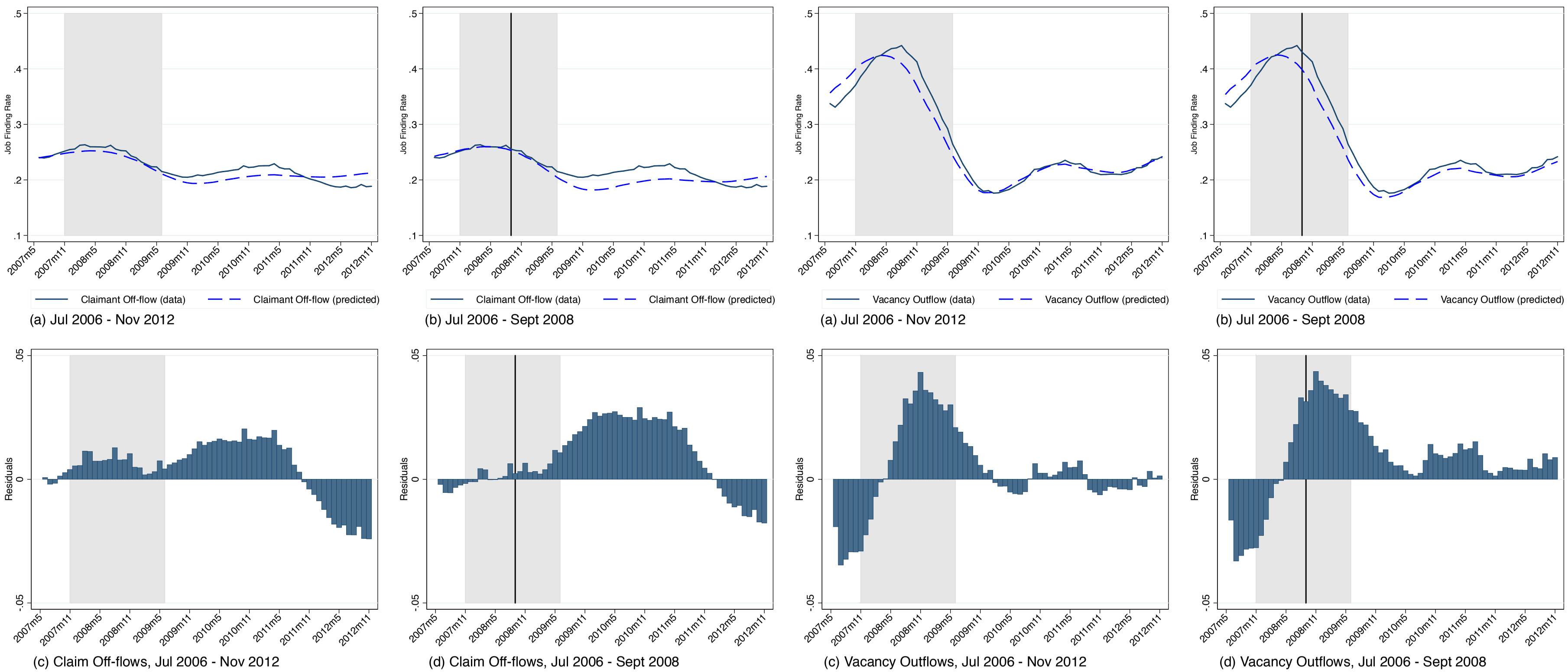


Figure A.6: Job Finding Rate: The Aggregate Matching Function (Nomis-OLS)

Note: The aggregate matching function estimated on the whole period and before the bankruptcy of Lehman Brothers. All series are 12-month moving averages. Shaded areas represent 2008-2009 recession as indicated by FRED (OECD).

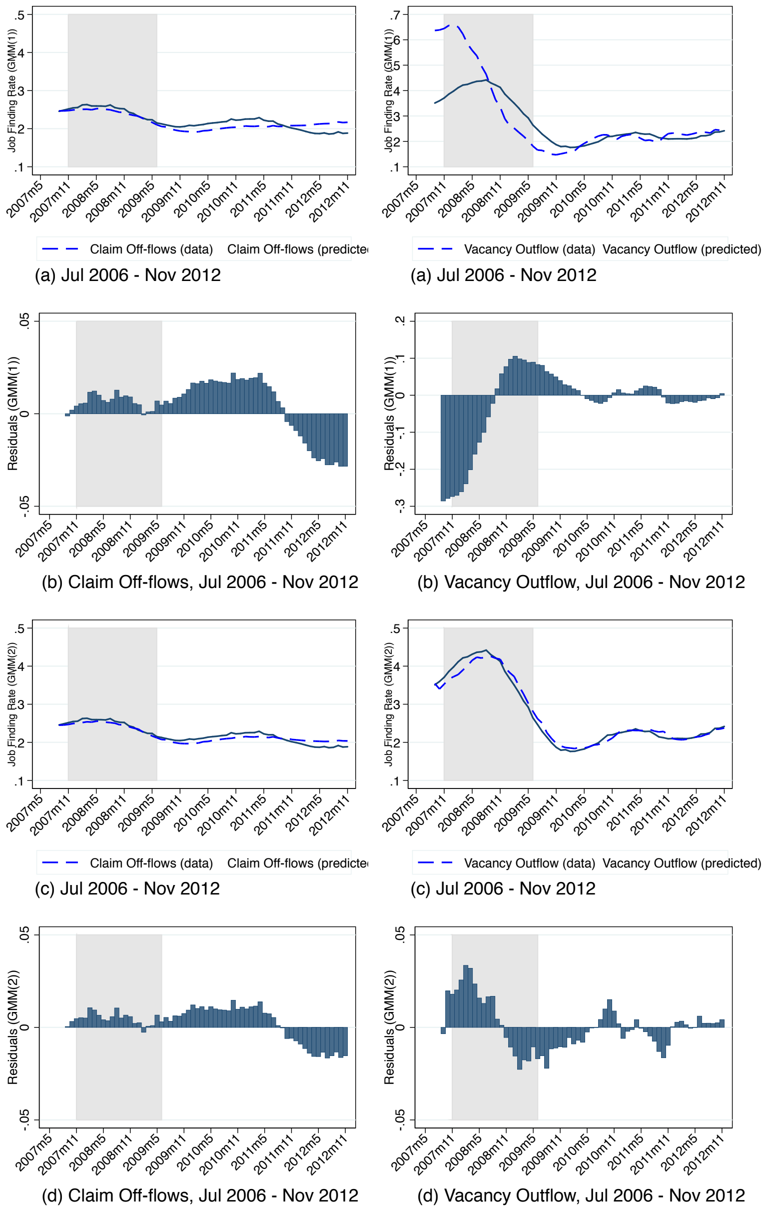


Figure A.7: Job Finding Rate: The Aggregate Matching Function (Nomis-GMM)

Note: The aggregate matching function estimated by GMM (Borowczyk-Martins et al. (2013)) on the whole period and before the bankruptcy of Lehman Brothers. All series are 12-month moving averages. Shaded areas represent 2008-2009 recession as indicated by FRED (OECD).

Table A.3: The Aggregate Matching Function: Elasticities

		Nomis		Longitudinal LFS	
		(1) Claim Off-fows	(2) Vacancy Outflows	(1) Short tenure	(2) UE transitions
OLS	$1-\sigma$	0.214**	0.709**	0.574**	0.337**
		(0.032)	(0.038)	(0.032)	(0.053)
	R^2	0.5201	0.8766	0.8941	0.8217
	test for CRS	p=0.03	p=0.21	p=0.10	p=0.33
	sample size	77	77	53	32
FD	$1-\sigma$	0.733**	0.986**	0.740**	0.140
		(0.279)	(0.285)	(0.200)	(0.053)
	R^2	0.3802	0.3514	0.8265	0.8665
	sample size	76	76	52	31
CES	$1-\sigma$	0.679	0.678**	0.815**	0.337
		(0.432)	(0.220)	(0.137)	(0.357)
	R^2	0.9861	0.9839	0.9962	0.9957
	sample size	77	77	53	32
GMM (1)	ARMA	(3,3)	(3,4)	(1,1)	–
	$1-\sigma$	0.213**	0.759**	0.593**	–
		(0.030)	(0.046)	(0.059)	–
	sample size	73	72	51	–
GMM (2)	ARMA	(3,3)	(3,4)	(1,1)	–
	$1-\sigma$	0.213**	0.688**	0.588**	–
		(0.051)	(0.218)	(0.035)	–
	Sargan test	1: 0.181: 0.671	1: 0.793: 0.373	1: 0.026: 0.872	–
	sample size	73	72	51	–
	frequency	monthly	monthly	quarterly	quarterly

Note: Standard errors in parentheses (robust for OLS and CES regressions). **Regression 1** uses Nomis unemployment and vacancy data, and claimant off-flows as a measure of matches. **Regression 2** uses Nomis unemployment and vacancy data, and vacancy outflows as a measure of matches. **Regression 3** uses LFS and VS data, and panel short tenure matches. **Regression 4** uses LFS and VS data, and panel UE transitions as matches. GMM (1) - just identified. GMM(2) - overidentified.

** significant at the 5 percent level. * significant at the 10 percent level.

Table A.4: Occupation (1-digit SOC2000) Segmented Matching Function: Elasticities (Nomis)

SIC2007	Specification		(1) Claim Off-fows	(2) Vacancy Outflows
whole period estimation				
Managers and Senior Officials		$1-\sigma$	0.159**	0.742**
	OLS		(0.029)	(0.042)
		R^2	0.5551	0.8635
		$1-\sigma$	0.063	0.500
	OLS (FD)		(0.230)	(0.296)
		R^2	0.2846	0.2340
Professional Occupations		$1-\sigma$	0.152**	0.741**
	OLS		(0.037)	(0.047)
		R^2	0.5660	0.8293
		$1-\sigma$	0.094	0.210
	OLS (FD)		(0.208)	(0.294)
		R^2	0.3559	0.3323
Associate Professional and Technical Occupations		$1-\sigma$	0.206**	0.687**
	OLS		(0.049)	(0.072)
		R^2	0.4982	0.6818
		$1-\sigma$	0.409**	0.476
	OLS (FD)		(0.197)	(0.265)
		R^2	0.3393	0.2697
Administrative and Secretarial Occupations		$1-\sigma$	0.166**	0.779**
	OLS		(0.022)	(0.040)
		R^2	0.6283	0.8409
		$1-\sigma$	0.217	0.051
	OLS (FD)		(0.119)	(0.164)
		R^2	0.3648	0.1867
Skilled Trades Occupations		$1-\sigma$	0.096**	0.743**
	OLS		(0.022)	(0.028)
		R^2	0.5808	0.9321
		$1-\sigma$	0.298	0.511*
	OLS (FD)		(0.226)	(0.262)
		R^2	0.4105	0.4286
Personal Service Occupations		$1-\sigma$	0.317**	0.694**
	OLS		(0.044)	(0.047)
		R^2	0.5330	0.7018
		$1-\sigma$	0.748**	0.890**
	OLS (FD)		(0.321)	(0.377)
		R^2	0.3658	0.2677
Sales and Customer Service occupations		$1-\sigma$	0.212**	0.743**
	OLS		(0.022)	(0.036)
		R^2	0.7116	0.8643
		$1-\sigma$	0.173	-0.018
	OLS (FD)		(0.168)	(0.169)
		R^2	0.3316	0.2970
Process, Plant and Machine Operatives		$1-\sigma$	0.115**	0.725**
	OLS		(0.025)	(0.032)
		R^2	0.5159	0.9116
		$1-\sigma$	0.168	0.476**
	OLS (FD)		(0.185)	(0.221)
		R^2	0.3997	0.3983
Elementary Occupations		$1-\sigma$	0.223**	0.616**
	OLS		(0.031)	(0.037)
		R^2	0.5629	0.8591
		$1-\sigma$	0.503**	0.632**
	OLS (FD)		(0.235)	(0.282)
		R^2	0.3545	0.3429
		sample size	77 (FD-76)	77 (FD-76)

Note: Robust standard errors in parentheses. **Regression 1** uses Nomis unemployment and vacancy data, and claimant off-flows as a measure of matches. **Regression 2** uses Nomis unemployment and vacancy data, and vacancy outflows as a measure of matches. ** significant at the 5 percent level. * significant at the 10 percent level.

Table A.5: OLS Fixed Effects

		Nomis		Longitudinal LFS	
		(1) Claim Off-fows	(2) Vacancy Outflows	(3) Short tenure	(4) UE transitions
whole period estimation					
Occupation	$1-\sigma$	0.173**	0.708**	-	-
		(0.021)	(0.028)	-	-
	R^2				
	within	0.5362	0.8395	-	-
	between	0.0219	0.9498	-	-
	overall	0.3161	0.8692	-	-
	sample size	9x77	9x77	-	-
Industry	$1-\sigma$	-	-	0.537**	0.317**
		-	-	(0.035)	(0.031)
	R^2				
	within	-	-	0.6489	0.5093
	between	-	-	0.3258	0.2960
	overall	-	-	0.3899	0.3157
	sample size	-	-	8x53	8x53
before the bankruptcy of Lehman Brothers					
Occupation	$1-\sigma$	0.165**	0.635**	-	-
		(0.036)	(0.094)	-	-
	R^2				
	within	0.7613	0.6011	-	-
	between	0.0244	0.8519	-	-
	overall	0.5554	0.6987	-	-
	sample size	9x26	9x26	-	-
Industry	$1-\sigma$	-	-	0.559**	0.063
		-	-	(0.179)	(0.089)
	R^2				
	within	-	-	0.5111	0.3215
	between	-	-	0.2368	0.2658
	overall	-	-	0.2235	0.1652
	sample size	-	-	8x25	8x25

Note: Robust standard errors in parentheses. **Regression 1** uses Nomis unemployment and vacancy data, and claimant off-flows as a measure of matches. **Regression 2** uses Nomis unemployment and vacancy data, and vacancy outflows as a measure of matches. **Regression 3** uses LFS and VS data, and panel short tenure matches. **Regression 4** uses LFS and VS data, and panel UE transitions as matches. All regressions are weighted by the average unemployment in each industry. ** significant at the 5 percent level. * significant at the 10 percent level.

Table A.6: Worker composition: by industry

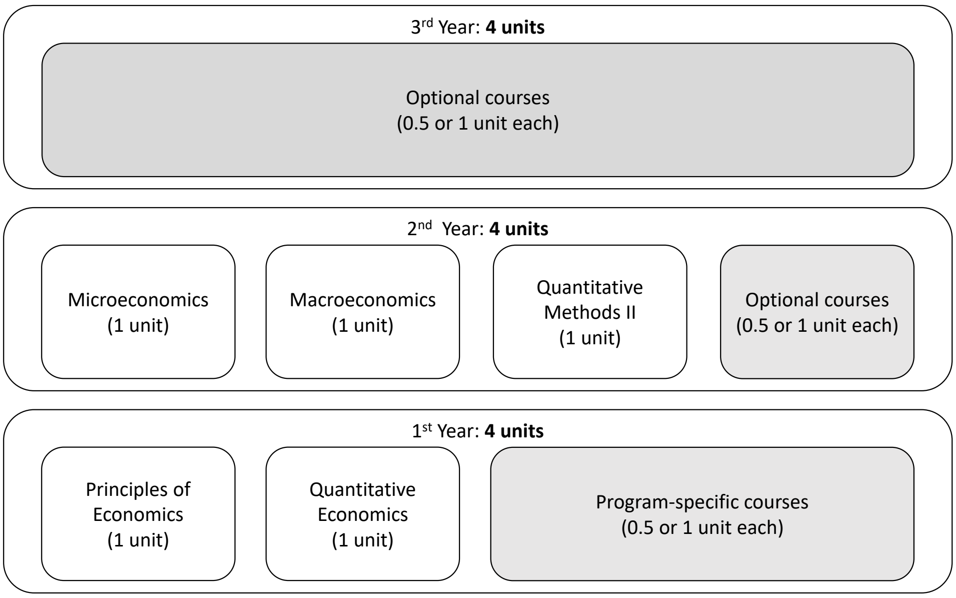
			Industry							
		Whole Sample	Energy & Water	Manufacturing	Construction	Hotels	Transport	Banking	Admin.	Other
No. of obs.		59,201	860	8,171	5,121	16,812	5,417	8,601	10,800	3,419
Age	<25	0.265	0.216	0.191	0.231	0.410	0.194	0.210	0.173	0.335
	>=25 & <50	0.519	0.500	0.519	0.524	0.451	0.548	0.572	0.579	0.487
	50+	0.216	0.284	0.290	0.245	0.139	0.258	0.219	0.248	0.178
U. dur.	< 6 months	0.603	0.621	0.544	0.538	0.616	0.575	0.635	0.640	0.625
	> 6 months	0.397	0.379	0.456	0.462	0.384	0.425	0.365	0.360	0.375
Ethnicity	White	0.870	0.917	0.888	0.933	0.853	0.858	0.852	0.861	0.889
	Black	0.038	0.027	0.027	0.023	0.032	0.045	0.049	0.049	0.039
	Asian	0.060	0.037	0.061	0.026	0.077	0.071	0.061	0.053	0.040
	Other	0.033	0.017	0.025	0.019	0.037	0.027	0.038	0.037	0.032
Education	Degree	0.153	0.156	0.094	0.053	0.074	0.163	0.250	0.280	0.178
	High Edu.	0.065	0.084	0.060	0.037	0.046	0.064	0.730	0.102	0.068
	A Level	0.184	0.187	0.161	0.197	0.199	0.168	0.188	0.177	0.187
	GCSE	0.302	0.283	0.296	0.357	0.349	0.290	0.260	0.243	0.311
	Other	0.048	0.034	0.060	0.055	0.058	0.047	0.036	0.037	0.040
	No Qual.	0.247	0.257	0.329	0.301	0.274	0.268	0.193	0.161	0.216
Imm. status	Native	0.863	0.903	0.869	0.921	0.862	0.858	0.838	0.851	0.874
	Immigrant	0.137	0.097	0.131	0.079	0.138	0.142	0.162	0.149	0.126
Region	North	0.268	0.288	0.305	0.292	0.268	0.252	0.233	0.268	0.249
	Midlands	0.175	0.179	0.234	0.154	0.176	0.184	0.146	0.168	0.141
	South	0.416	0.315	0.315	0.381	0.417	0.437	0.489	0.428	0.466
	Wales	0.047	0.056	0.061	0.059	0.045	0.040	0.040	0.051	0.047
	Scotland	0.093	0.162	0.084	0.114	0.094	0.087	0.092	0.084	0.097
Dep. child.	No Child.	0.622	0.713	0.675	0.673	0.585	0.658	0.641	0.575	0.619
	1	0.191	0.158	0.160	0.159	0.222	0.0154	0.180	0.206	0.214
	2	0.128	0.084	0.109	0.110	0.131	0.127	0.122	0.158	0.117
	3+	0.059	0.045	0.056	0.057	0.063	0.061	0.057	0.061	0.050
Gender	Male	0.575	0.813	0.771	0.923	0.497	0.755	0.547	0.315	0.514
	Female	0.425	0.187	0.229	0.077	0.503	0.245	0.453	0.685	0.486
Whole sample			0.015	0.138	0.087	0.284	0.092	0.145	0.182	0.058

Note: Sample period 2001 Q3 – 2014 Q3.

Appendix B

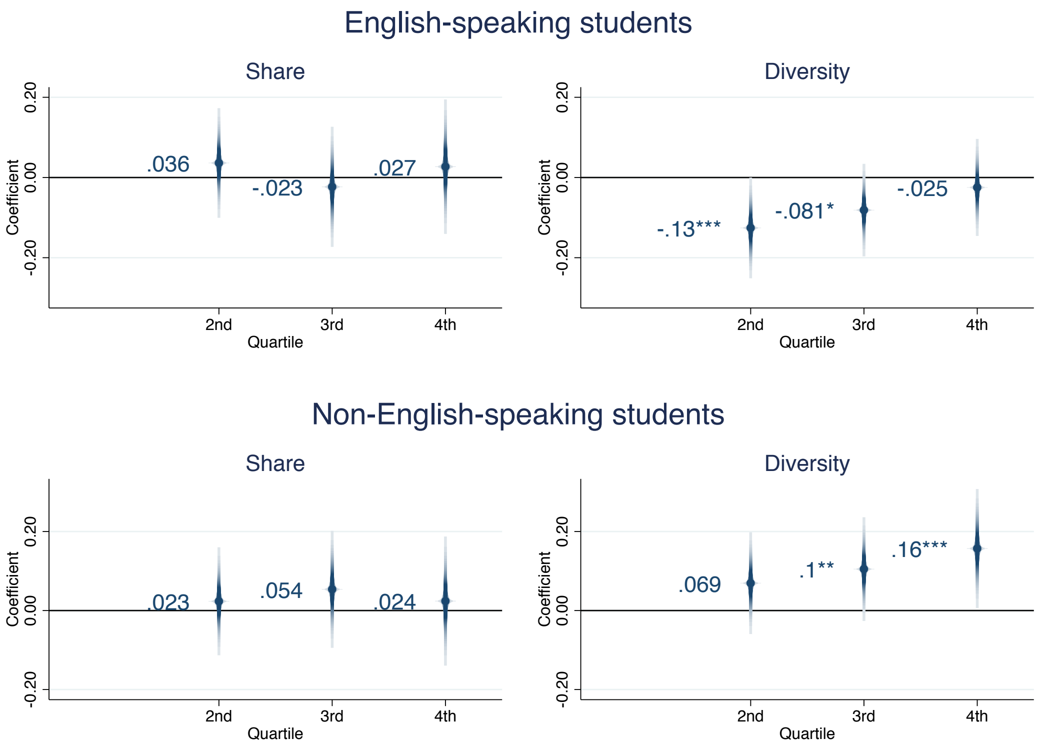
Chapter 2 Appendix: Peer Diversity, College Performance and Educational Choices

Figure B.1: Structure of teaching



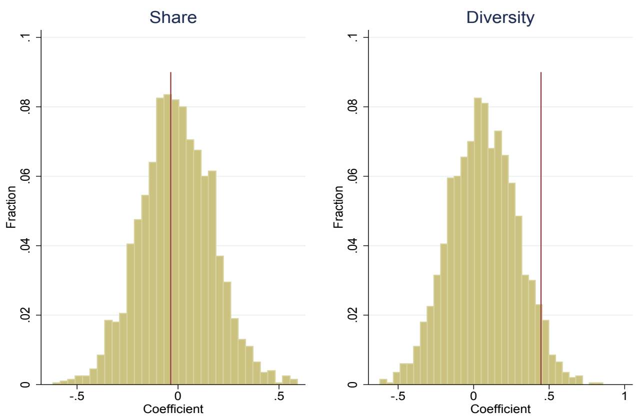
Notes: This figure describes the teaching structure of the institutional setting. Teaching happens in three consecutive years. Per year, students take four teaching units. In our specifications, we rely on quasi-random assignment into seminars within first- and second-year courses. Non-compulsory optional courses (grey) are not part of our sample. Third-year course choices are regarded as outcomes in Section 2.4.4.

Figure B.2: Non-linear effects of Share and Diversity on Contemporaneous Grades



Notes: This figure shows the estimates of the effect of non-English-speaking share on grades, when share of non-English-speaking students is approximated by dummies indicating its quartile in the overall distribution of non-English-speaking students shares.

Figure B.3: Distribution of placebo estimates



Notes: This figure shows the empirical distributions of placebo estimates for the share of non-English-speaking students (left) and ethno-linguistic diversity among those (right) on course grades. The cumulative distribution functions are based on 2000 estimates using a specification similar to the one displayed in column (2) of Table 2.3 lower panel and using random permutations of seminar ID to compute the treatments. The vertical line indicates the original estimate. Implied p-values are 0.556 (left) and 0.044 (right).

Table B.1: Sample composition by language background

Language	Associated Nationalities	Number of Speakers	Share in sample (%)
ENGLISH	United States Ireland Australia New Zealand Kenya Uganda United Kingdom British Indian Ocean Territory British Overseas Citizen Nigeria Trinidad & Tobago Gambia Canada South Africa	971	44.46
MANDARIN	China Singapore Taiwan	420	19.23
RUSSIAN	Russia Kazakhstan	106	4.85
ITALIAN	Italy	68	3.11
CANTONESE	Hong Kong Macao	42	1.92
FRENCH	France	41	1.88
BULGARIAN	Bulgaria	39	1.79
KOREAN	North Korea South Korea	38	1.74
GERMAN	Germany Austria	33	1.51
POLISH	Poland	26	1.19
ARABIC	Bahrain Saudi Arabia Lebanon United Arab Emirates Libya Oman Morocco Kuwait Egypt Jordan Algeria	24	1.11
GREEK	Greece Cyprus	23	1.05
VIETNAMESE	Vietnam	21	0.96
SWEDISH	Sweden	20	0.92
PORTUGESE	Portugal Brazil Angola	18	0.82
SPANISH	Spain Mexico Columbia El Salvador	18	0.82
LITHUANIAN	Lithuania	17	0.78
WESTERN PUNJABI	Pakistan	17	0.78
AZERBAIJANI	Azerbaijan	17	0.78
HINDI	India	16	0.73
ALL OTHER (48)		209	9.57
Total Sample		2184	

Notes: This table gives the number of individual speakers of top 20 languages as well as the share of that particular language in our full sample.

Table B.2: Raw and residual variation in key variables

	Mean	*SD*	*Min*	*Max*
Absolute				
Share of non-English speakers	0.54	0.14	0.21	0.87
Blau index of diversity	0.87	0.10	0.29	1.00
Residualised				
Share of non-English speakers	0.00	0.09	-0.29	0.33
Blau index of diversity	0.00	0.07	-0.43	0.23
No. of obs	*8744*			

Notes: This table shows variation in the share of non-English speakers and the diversity index, in absolute levels and in residualised after controlling for course×year, study programme, day×hour, and seminar leader fixed effects.

Table B.3: Robustness: Diversity and educational performance, controlling for ability

Sample	Grade	Fail	Honour
Total			
Share of non-English	-0.098	0.087***	0.011
	(0.094)	(0.033)	(0.049)
Blau Index	0.234***	-0.123***	-0.033
	(0.088)	(0.040)	(0.049)
Mean of dep. var.	0.000	0.167	0.401
R^2	0.48	0.25	0.36
No. of observations	8680	8680	8680
English			
Share of non-English	-0.099	0.080	-0.047
	(0.138)	(0.053)	(0.067)
Blau Index	0.117	-0.106*	0.049
	(0.121)	(0.057)	(0.071)
Mean of dep. var.	0.079	0.146	0.427
R^2	0.46	0.23	0.36
No. of observations	4003	4003	4003
Non-English			
Share of non-English	-0.059	0.094*	0.060
	(0.131)	(0.050)	(0.066)
Blau Index	0.345***	-0.135**	-0.082
	(0.133)	(0.058)	(0.065)
Mean of dep. var.	-0.068	0.185	0.379
R^2	0.52	0.28	0.38
No. of observations	4677	4677	4677
Course × year FE	yes	yes	yes
Study program FE	yes	yes	yes
Day/Time FE	yes	yes	yes
Seminar leader FE	yes	yes	yes
Seminar controls	yes	yes	yes
Individual controls	yes	yes	yes
Ability control	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables (standardised grade, indicator for failing a course, indicator for receiving an honour (60% or above)) on the seminar-wise leave-me-out share of non-English speakers and the diversity index. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. Seminar controls are share of females, number of students and mean age. In addition, here we control for students' and their peers' GPA. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table B.4: Robustness: Coefficient stability to seminar controls

Sample	without ability controls		with ability controls	
	without Seminar Controls	with Seminar Controls	without Seminar Controls	with Seminar Controls
Total				
Share of non-English	-0.036	-0.094	-0.047	-0.098
	(0.116)	(0.129)	(0.086)	(0.094)
Blau Index	0.249*	0.275**	0.207**	0.234***
	(0.138)	(0.139)	(0.093)	(0.088)
Mean of dep. var.	0.000	0.000	0.000	0.000
R^2	0.05	0.05	0.48	0.48
No. of observations	8744	8744	8680	8680
English				
Share of non-English	0.094	-0.033	0.028	-0.099
	(0.163)	(0.175)	(0.129)	(0.138)
Blau Index	-0.075	0.004	0.032	0.117
	(0.159)	(0.162)	(0.124)	(0.121)
Mean of dep. var.	0.079	0.079	0.079	0.079
R^2	0.08	0.08	0.46	0.46
No. of observations	4032	4032	4003	4003
Non-English				
Share of non-English	-0.142	-0.099	-0.082	-0.059
	(0.165)	(0.177)	(0.119)	(0.138)
Blau Index	0.515***	0.476**	0.369***	0.345***
	(0.197)	(0.201)	(0.130)	(0.133)
Mean of dep. var.	-0.068	-0.068	-0.068	-0.068
R^2	0.08	0.08	0.52	0.52
No. of observations	4712	4712	4677	4677
Course × year FE	yes	yes	yes	yes
Study program FE	yes	yes	yes	yes
Day/Time FE	yes	yes	yes	yes
Seminar leader FE	yes	yes	yes	yes
Individual controls	yes	yes	yes	yes
Seminar controls	no	yes	no	yes
Individual ability	no	no	yes	yes
Peer ability	no	no	no	yes

Notes: This table summarises results of regressions of standardised grades on the seminar-wise leave-me-out share of non-native speakers and the diversity index. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. Seminar controls are share of females, number of students and mean age. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table B.5: Robustness: Alternative language group definitions

Sample	Baseline	Predominant languages	Nationality
Total			
Share of non-English	-0.094	-0.106	-0.109
	(0.129)	(0.126)	(0.126)
Blau Index	0.275**	0.237	0.337**
	(0.139)	(0.147)	(0.163)
Mean of dep. var.	0.000	0.000	0.000
R^2	0.05	0.05	0.05
No. of observations	8744	8744	8744
English			
Share of non-English	-0.033	-0.076	-0.098
	(0.175)	(0.173)	(0.177)
Blau Index	0.004	0.049	0.114
	(0.162)	(0.176)	(0.203)
Mean of dep. var.	0.079	0.084	0.100
R^2	0.08	0.08	0.08
No. of observations	4032	3970	3768
Non-English			
Share of non-English	-0.099	-0.097	-0.076
	(0.177)	(0.176)	(0.169)
Blau Index	0.476**	0.384*	0.496**
	(0.201)	(0.201)	(0.208)
Mean of dep. var.	-0.068	-0.070	-0.076
R^2	0.08	0.08	0.08
No. of observations	4712	4774	4976
Course × year FE	yes	yes	yes
Study program FE	yes	yes	yes
Day/Time FE	yes	yes	yes
Seminar leader FE	yes	yes	yes
Seminar controls	yes	yes	yes
Individual controls	yes	yes	yes

Notes: This table summarises results of regressions of standardised grades on the seminar-wise leave-me-out share of non-native speakers and different definitions of the diversity index. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. In column 2, students are given the predominant language of their country, and are not considered native speakers even if English is an official (but not-predominant) language. In column 3, only the UK nationals are considered to be native speakers. Seminar controls are share of females, number of students and mean age. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table B.6: Robustness: Alternative diversity definitions

Sample	Baseline	No. of languages	No. of same lang.	Share of same lang.	At least one of same lang.
Total					
Share of non-English	-0.094	-0.332**	–	–	–
	(0.129)	(0.161)	–	–	–
Diversity	0.275**	0.022***	–	–	–
	(0.139)	(0.007)	–	–	–
Mean of dep. var.	0.000	0.000	–	–	–
R^2	0.05	0.05	–	–	–
No. of observations	8744	8744	–	–	–
English					
Share of non-English	-0.033	-0.160	–	–	–
	(0.175)	(0.211)	–	–	–
Diversity	0.004	0.011	–	–	–
	(0.162)	(0.009)	–	–	–
Mean of dep. var.	0.079	0.079	–	–	–
R^2	0.08	0.08	–	–	–
No. of observations	4032	4032	–	–	–
Non-English					
Share of non-English	-0.099	-0.415*	-0.062	-0.057	-0.068
	(0.177)	(0.218)	(0.179)	(0.179)	(0.179)
Diversity	0.476**	0.029***	-0.013*	-0.452**	-0.090***
	(0.201)	(0.009)	(0.007)	(0.197)	(0.032)
Mean of dep. var.	-0.068	-0.068	-0.068	-0.068	-0.068
R^2	0.08	0.08	0.07	0.07	0.07
No. of observations	4712	4712	4712	4712	4712
Course × year FE	yes	yes	yes	yes	yes
Study program FE	yes	yes	yes	yes	yes
Day/Time FE	yes	yes	yes	yes	yes
Seminar leader FE	yes	yes	yes	yes	yes
Seminar controls	yes	yes	yes	yes	yes
Individual controls	yes	yes	yes	yes	yes

Notes: This table summarises results of regressions of standardised grades on the seminar-wise leave-me-out share of non-native speakers and different definitions of the diversity index. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. In column 2, students are given the predominant language of their country, and are not considered native speakers even if English is an official (but not-predominant) language. In column 3, only the UK nationals are considered to be native speakers. Seminar controls are share of females, number of students and mean age. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table B.7: The role of Mandarin speakers

	Share of Mandarin Speakers	Baseline	inc. Nbr Mandarin Speakers	inc. Share Mandarin Speakers	Mandarin Speakers only	Other Speakers only
Non-English						
Share of non-English	0.283***	-0.099	-0.182	-0.110	-0.188	-0.170
	(0.023)	(0.177)	(0.205)	(0.210)	(0.238)	(0.345)
Blau Index	-0.633***	0.476**	0.669**	0.511	0.288	0.695
	(0.029)	(0.201)	(0.338)	(0.337)	(0.444)	(0.606)
R^2	0.86	0.08	0.08	0.08	0.10	0.19
No. of observations	8744	4712	4712	4712	3202	1509

Notes: This table summarises results of different robustness checks examining the role of Chinese Mandarin speakers. Specification (1) explores the relationship between the share of Mandarin speakers and total share of non-native English speakers as well as diversity. Specification (2) lists the baseline results similar to Table 2.3, column (2). Column (3) displays results controlling for the number of Chinese students in the seminar. Column (4) displays results controlling for the share of Chinese students. Columns (5) and (6) repeat this specification separately for Chinese and other non-English speakers. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. Seminar controls are share of females, number of students and mean age. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table B.8: Robustness of inference

	Baseline	i.i.d.	robust	course/year	year
English					
Share of non-English	-0.033	-0.033	-0.033	-0.033	-0.033
	(0.175)	(0.181)	(0.185)	(0.172)	(0.131)
Blau Index	0.004	0.004	0.004	0.004	0.004
	(0.162)	(0.211)	(0.217)	(0.192)	(0.208)
R^2	0.08	0.08	0.08	0.08	0.08
No. of observations	4032	4032	4032	4032	4032
Non-English					
Share of non-English	-0.099	-0.099	-0.099	-0.099	-0.099
	(0.177)	(0.174)	(0.172)	(0.211)	(0.310)
Blau Index	0.476**	0.476**	0.476**	0.476**	0.476**
	(0.201)	(0.208)	(0.210)	(0.207)	(0.188)
R^2	0.08	0.08	0.08	0.08	0.08
No. of observations	4712	4712	4712	4712	4712

Notes: This table summarises results of different robustness checks on inference. Specification (1) displays the baseline specification similar to Table 2.3, column (2) with standard errors clustered at the seminar level. Column (2) lists results assuming i.i.d. error terms. Column (3) lists results based on robust standard errors. Column (4) displays standard errors clustered on the course×year level. Column (5) applies standard errors clustered on the year level. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. Seminar controls are share of females, number of students and mean age. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.9: Questions of the field survey

Question
How do you rate your English proficiency? **[Very bad (1) - Very good (5)]**
How comfortable do you feel speaking in English in this tutorial? **[Very uncomfortable (1) - Very comfortable (5)]**
For this course how often do you work with: Native English-speaking students **[Never (1) - Very often (5)]**
For this course how often do you work with: Non-native English-speaking students **[Never (1) - Very often (5)]**
How do you rate the level of English in the seminar discussions? **[Very bad (1) - Very good (5)]**